

EVOLUÇÃO DE REDES
PONDERADAS E CONEXAS:
ANÁLISE DE UMA REDE
DE PREFERÊNCIAS MUSICAIS

POR

EDUARDA MARIA PORTELA GUIMARÃES DE SOUSA

TESE DE MESTRADO EM MODELAÇÃO, ANÁLISE DE DADOS
E SISTEMAS DE APOIO À DECISÃO

SETEMBRO DE 2014

ORIENTADA POR:

PROF. PEDRO CAMPOS

PROF.^a PAULA BRITO

Faculdade de Economia

Universidade do Porto

2014

Palavras chave: Redes sociais, Redes ponderadas, *Blockmodeling*, Dinâmica de Redes

Breve nota biográfica

Eduarda Maria Portela Guimarães de Sousa nasceu em Vila Nova de Famalicão em Junho de 1981. Desde então reside na Trofa.

Após conclusão do 12º Ano na Escola Secundária da Trofa ingressou na Licenciatura em Matemática Aplicada à Tecnologia na Faculdade de Ciências da Universidade do Porto. Depois de alguns anos afastada acabou por retomar os estudos e concluir a Licenciatura em Matemática Aplicada, uma vez que entretanto houve uma reestruturação pelo processo de Bolonha. Desde então trabalha como explicadora de Matemática em diferentes Centros de Estudo perto da sua zona de residência.

No intuito de aprofundar os conhecimentos entrou em 2011 para a Faculdade de Economia da Universidade do Porto no Mestrado de Modelação, Análise de Dados e Sistemas de Apoio à Decisão. Presentemente concilia a atividade profissional com a conclusão da tese de mestrado na área de análise de redes sociais.

Agradecimentos

No decorrer deste trabalho deparei-me com situações que não seriam ultrapassadas sem o apoio incondicional de pessoas muito especiais.

Começo por agradecer aos meus Orientadores, Prof. Pedro Campos e Prof.^a Paula Brito por toda a motivação e disponibilidade. Foram incansáveis na orientação deste trabalho até ao último minuto.

À UBBIN pela disponibilização dos dados analisados neste trabalho.

À minha família, especialmente a minha mãe, Maria Antonieta e à minha madrinha, Maria Eduarda.

À Maria, ao Alexandre, ao Rui e à Lidiane, colegas de curso e grandes amigos que prestaram um apoio incondicional.

À Daniela e demais amigos de longa data, por todo o ânimo que me deram neste percurso.

Por fim, ao Nelson que está sempre do meu lado.

A todos um meu Muito Obrigada!

Resumo

O objetivo deste trabalho é estudar metodologias de análise de dinâmica de redes sociais no intuito de as aplicar a uma rede de preferências musicais. Cada vértice desta rede representa um estilo de música. Cada aresta é ponderada; a ponderação de uma aresta representa o número de pessoas que, nas suas preferências, indicam em simultâneo os estilos de música referentes aos vértices que esta une.

Para este estudo é proposta uma análise das métricas específicas de análise de redes ponderadas tanto a nível de rede como a nível de vértice.

Estas medidas são combinadas com as perspetivas janela deslizante e janela acumulada no intuito de estudar a dinâmica de rede.

Para além desta análise são também aplicados algoritmos de deteção de comunidades, com especial destaque para o *Blockmodeling*, que tem como objetivo subdividir a rede em grupos de vértices que se ligam da mesma forma entre si e entre os restantes vértices na rede.

Por fim será estudada a dinâmica das comunidades formadas por um método de deteção de eventos.

Abstract

The objective of this work is the study of methodologies for analysis of networks' dynamics in order to apply them to a musical preferences network.

Each vertex of this network represents a style of music. Each edge is weighted; the weight of each edge is the number of people who chose simultaneous for their preference list the two styles of music that are connected by the edge.

An analysis of specific metrics associated with valued networks, both at vertex and network level, is proposed in this work. These metrics are combined with prospects like sliding window and accumulated widow so as to analyze the networks dynamics.

Besides the analyses mentioned above we will also apply community detection algorithms, with special emphasis to Blockmodeling which aims at subdividing the set of vertices into groups that are connected in the same way between themselves and the rest of the network vertices.

Finally, the dynamics of the obtained communities will be studied using methods of event detection.

Índice

<i>Breve nota biográfica</i>	<i>ii</i>
<i>Agradecimentos</i>	<i>iii</i>
<i>Resumo</i>	<i>iv</i>
<i>Abstract</i>	<i>v</i>
<i>Índice de Figuras</i>	<i>ix</i>
<i>Índice de Tabelas</i>	<i>xi</i>
<i>Capítulo 1</i>	<i>1</i>
<i>Introdução</i>	<i>1</i>
<i>1.1. Motivação</i>	<i>1</i>
<i>1.2. Objetivos</i>	<i>2</i>
<i>1.3. Contribuições</i>	<i>2</i>
<i>1.4. Organização</i>	<i>3</i>
<i>Capítulo 2</i>	<i>4</i>
<i>Análise de Redes Sociais</i>	<i>4</i>
<i>2.1. Teoria de Grafos</i>	<i>5</i>
<i>2.1.1. Conceitos sobre Grafos</i>	<i>6</i>
<i>2.1.2. Grafos direcionados (Digrafos)</i>	<i>7</i>
<i>2.1.3. Grafos Ponderados</i>	<i>8</i>
<i>2.1.4. Grafos completos, nulos e regulares</i>	<i>10</i>
<i>2.1.5. Árvores e Grafos Bipartidos</i>	<i>11</i>
<i>2.2. Medidas Estatísticas de Análise de Redes</i>	<i>12</i>
<i>2.2.1. Medidas a nível dos vértices (atores)</i>	<i>13</i>
<i>2.2.1.1. Excentricidade</i>	<i>13</i>
<i>2.2.1.2. Grau</i>	<i>13</i>
<i>2.2.1.3. Intermediação (Betweenness) de um vértice</i>	<i>14</i>
<i>2.2.1.4. Proximidade</i>	<i>16</i>
<i>2.2.1.5. Coeficiente de Clustering Local</i>	<i>17</i>
<i>2.2.2. Medidas a nível de Rede</i>	<i>19</i>
<i>2.2.2.1. Raio e Diâmetro duma rede</i>	<i>19</i>

2.2.2.2. <i>Densidade da Rede</i>	19
2.2.2.3. <i>Distância Geodésica Média</i>	20
2.2.2.4. <i>Grau Médio</i>	20
2.2.2.5. <i>Coefficiente de Clustering Global</i>	21
2.3. <i>Estudo da Dinâmica de Redes</i>	21
Capítulo 3	23
<i>Deteção de Comunidades</i>	23
3.1. <i>Blockmodeling</i>	23
3.1.1. <i>Bloco</i>	25
3.1.2. <i>Tipos de relações</i>	25
3.1.2.1. <i>Relação Completa</i>	25
3.1.2.2. <i>Relação Regular</i>	26
3.1.2.3. <i>Relação Nula</i>	26
3.1.3. <i>Tipos de Equivalência</i>	26
3.1.4. <i>Blockmodel</i>	27
3.1.5. <i>Definição Formal</i>	29
3.1.6. <i>Generalização de Blockmodeling a Redes Ponderadas</i>	31
3.2. <i>Clustering Hierárquico</i>	32
3.3. <i>Modularidade</i>	34
3.3.1. <i>Algoritmo CNM</i>	35
3.3.2. <i>Walktrap</i>	35
3.3.3. <i>Algoritmo de Girvan and Newman</i>	36
3.4. <i>Evolução de Redes - Deteção de Eventos</i>	36
3.4.1. <i>Algoritmo de Choobdar, et al</i>	37
3.4.2. <i>MONIC</i>	38
Capítulo 4	40
<i>Caso de estudo: Palco Principal</i>	40
4.1. <i>O que é o Palco Principal</i>	40
4.2. <i>Informação disponível para o estudo</i>	41
4.3. <i>Metodologia aplicada na elaboração deste estudo</i>	44
4.4. <i>Extração de dados a partir da informação disponível</i>	47
4.5. <i>Estudo das matrizes originais nos primeiro e quinto instantes</i>	49

4.5.1. Medidas a nível de Rede:.....	49
4.5.2. Medidas a nível dos vértices:.....	50
4.6. Janela Deslizante e Janela Acumulada para aparecimento/aumento de ligações	53
4.6.1. Janelas Deslizante	54
4.6.2. Janela Acumulada para aparecimento/aumento de ligações:.....	66
4.7. Detecção de Comunidades	69
4.7.1. Aplicação de otimização pelo algoritmo Blockmodeling	70
4.7.2. Detecção de Eventos	73
4.8. Aplicação do Blockmodeling à partição obtida por Clustering Hierárquico....	76
Capítulo 5.....	81
Conclusões	81
5.1. Trabalho futuro.....	82
Bibliografia.....	83
Anexo 1 – Tabelas com valores de medidas a nível de vértice em A1, A2 e A3 considerando $\alpha = 1.5$.....	85
Anexo 2 – Tabelas de contribuições de vértices entre as janelas D2 e D3 e representação em rede das mesmas janelas	86

Índice de Figuras

Figura 1 - Grafo simples	4
Figura 2 - Matriz de Adjacência do grafo da Figura 1	6
Figura 3 - Rede de fornecimento entre empresas (à esquerda) e Mapa com informação de sentido de ruas numa localidade (à direita).....	7
Figura 4 - Matriz de Adjacência do Mapa da Figura 3 (direita).....	8
Figura 5 – Exemplo de Grafo Ponderado.....	8
Figura 6 - Matriz de Adjacência do grafo direcionado e ponderado representado na Figura 4	9
Figura 7 – Grafo ponderado das horas que cinco indivíduos passam juntos por semana	9
Figura 8 – Grafo ponderado depois da transformação dos valores pelo método sugerido por Tore Opsahl	10
Figura 9 - Exemplo de grafo completo (em cima à esquerda), nulo (em cima à direita) e regular (em baixo).....	11
Figura 10 - Matriz de Adjacência do Grafo Regular da Figura 9 (em baixo)	11
Figura 11 – Grafo Bipartido.....	12
Figura 12 - Exemplo de um grafo	12
Figura 13 - Exemplo de <i>Gatekeepers</i>	15
Figura 14 - Exemplo de Ponte Local.....	16
Figura 15- Exemplo dos valores de proximidade entre vértices	17
Figura 16- Vizinhança do vértice A	18
Figura 17- Grafo com 6 vértices	20
Figura 18 – Janela Acumulada, W_j (em cima) com avanço r) e Janela Deslizante, W'_j (em baixo) com avanço r e dimensão d	22
Figura 19 – Partição de um grafo em 3 <i>clusters</i>	24
Figura 20- Matriz de Adjacência do Grafo da Figura 19	24
Figura 21 - Matriz de Adjacência de uma Rede.....	27
Figura 22- Matriz de Adjacência equivalente após permutações.....	28
Figura 23- Matriz reduzida da matriz da Figura 19	29
Figura 24- Diagrama reduzido do grafo da Figura 19.....	30
Figura 25 - Exemplo de Dendrograma da aplicação de <i>Clustering</i> Hierárquico usando a Distância Euclideana e o Critério do Valor Médio (USArrests - R database)	33
Figura 26 – Parte do <i>Interface</i> do site “Palco Principal” referente ao registo de preferências dos utilizadores (palcoprincipal.sapo.pt)	41
Figura 27 – Parte da matriz de preferências referentes a setembro de 2011	42
Figura 28 - Diagrama da rede referente a setembro de 2011 de onde retiraram os <i>loops</i>	43
Figura 29 – Esquema de processo de partição pelo algoritmo CNM e posterior otimização pelo algoritmo de <i>Blockmodeling</i>	45
Figura 30 - Esquema do processo de partição por <i>Clustering</i> Hierárquico e posterior otimização por <i>Blockmodeling</i>	46
Figura 31 - Operações efetuadas nas matrizes 01, 02, 03, 04, 05	47
Figura 32 - Construção das Janelas Deslizantes.....	48

Figura 33 - Construção das Janelas Acumuladas	48
Figura 34 – Representação em rede de D1	54
Figura 35 - Representação em rede de D2	55
Figura 36 - Representação em rede de D3	55
Figura 37 – Janela Acumulada A1	66
Figura 38 – Janela Acumulada A2	67
Figura 39 – <i>Clusters</i> do primeiro instante da Janela Deslizante D1 formados pelo método <i>Fastgreedy</i>	71
Figura 40 - <i>Clusters</i> formados pela aplicação de otimização por Blockmodeling a D1	71
Figura 41 - Blocos formados na matriz D1	72
Figura 42 - <i>Clusters</i> sobreviventes entre D1 e D2	74
Figura 43 – Nascimento de <i>clusters</i> entre D1 e D2	74
Figura 44 - <i>Clusters</i> sobreviventes entre D2 e D3	75
Figura 45 – Nascimento de <i>clusters</i> entre D2 e D3	75
Figura 46 – Dendrograma obtido pela aplicação de <i>clustering</i> hierárquico a D1	76
Figura 47 – <i>Clusters</i> formados pela aplicação de otimização por Blockmodeling a D1	77
Figura 48 – <i>Clusters</i> formados pela aplicação de otimização por Blockmodeling a D2	78
Figura 49 – <i>Clusters</i> formados pela aplicação de otimização por Blockmodeling a D3	78
Figura 50 – Segunda Janela Acumulada	79
Figura 51 – Partição do Grafo da janela D2 através da aplicação do <i>blockmodeling</i>	86
Figura 52 – Partição do Grafo da janela D3 através da aplicação do <i>blockmodeling</i>	87

Índice de Tabelas

Tabela 1 - Matriz de distâncias no grafo da Figura 7.....	13
Tabela 2 - Grau para diferentes valores de α do exemplo da Figura 7.....	14
Tabela 3 - Coeficiente de <i>Clustering</i> Local para o exemplo da Figura 7.....	19
Tabela 4 - Coeficiente de <i>Clustering</i> Global.....	21
Tabela 5 – Inconsistências de Bloco em <i>Homogeneity Blockmodeling</i>	31
Tabela 6 - Medidas de rede ao longo dos 5 Instantes.....	49
Tabela 7 – Coeficiente de <i>Clustering</i> Local ignorando as ponderações das arestas	51
Tabela 8 – Coeficiente de <i>Clustering</i> Local calculado pela Distância Geométrica	51
Tabela 9 - Grau para rede binária	Tabela 10 – Grau para rede ponderada.....
Tabela 11 – Cinco vértices que apresentam valores mais altos de Proximidade com $\alpha=0$ (tabela em cima à esquerda) e $\alpha=1.5$ (em baixo à esquerda) e de Intermediação com $\alpha=0$ (tabela em cima à direita) e $\alpha=1.5$ (em baixo à direita) para os instantes T1 e T5	52
Tabela 12 - Tabela com as medidas a nível de rede para aparecimento/fortalecimento no método Janela Deslizante referentes a D1, D2 e D3	56
Tabela 13 - Excentricidade dos vértices para ganho/fortalecimento de ligações em D1, D2 e D3	58
Tabela 14 – Vértices com maior e menor Grau na rede quando esta é considerada binária ($\alpha=0$).....	59
Tabela 15 - Vértices com maior e menor Grau na rede quando é dada mais importância à ponderação da aresta do que ao número de vizinhos ($\alpha=1.5$).....	59
Tabela 16 - <i>Closeness</i> quando a Rede é considerada binária	60
Tabela 17 - <i>Closeness</i> quando é dada maior importância à ponderação das arestas.....	61
Tabela 18 – Intermediação quando $\alpha=0$ (Rede binária).....	62
Tabela 19 – Intermediação quando $\alpha=1.5$ (Rede ponderada)	62
Tabela 20 – Vértices para os quais não é calculada a medida do Coeficiente de <i>Clustering</i> (em cima) e vértices para os quais esta medida é um (em baixo) para as janelas D1, D2 e D3	64
Tabela 21 - Valores de Coeficiente de <i>Clustering</i> supondo a matriz binária.....	64
Tabela 22 - Valores de Coeficiente de <i>Clustering</i> usando a Distância Geométrica.....	65
Tabela 23 – Medidas a nível de rede para A1, A2 e A3	67
Tabela 24 – Valores de Intermediação para a Janelas Acumulada quando $\alpha=1,5$	69
Tabela 25 – Distribuição dos <i>clusters</i> de D1 pelos <i>clusters</i> de D2	73
Tabela 26 – Contribuições de cada <i>cluster</i> de D1 para cada <i>cluster</i> de D2	73
Tabela 27 – Valores de Excentricidade em A1, A2 e A3 com $\alpha = 1.5$	85
Tabela 28 - Valores de Grau em A1, A2 e A3 com $\alpha = 1.5$	85
Tabela 29 - Valores de Proximidade em A1, A2 e A3 com $\alpha = 1.5$	85
Tabela 30 – Distribuição dos <i>clusters</i> de D2 pelos <i>clusters</i> de D3	86
Tabela 31 – Contribuições de cada <i>cluster</i> de D2 para cada <i>cluster</i> de D3	86

Capítulo 1

Introdução

A modelação da informação no formato de redes sociais foca-se na definição de relações (pessoais, profissionais, lúdicas) entre diferentes entidades (empresas ou indivíduos), e não apenas nas características individuais de cada um, permitindo conduzir estudos em relação à estrutura das redes. Esta perspetiva de análise permite a aplicação de novas metodologias que complementam ou até superam outros tipos de análise.

Esta temática tem sofrido evoluções durante os últimos anos, motivadas tanto pelo crescente volume de informação gerado mundialmente como pelas inovações tecnológicas que têm surgido.

Matematicamente o diagrama que representa tais situações é designado por grafo e é utilizado nas mais diversas áreas de conhecimento como por exemplo a Biologia na representação de ecossistemas, no campo do Jornalismo para estudar a velocidade de difusão de informação e nas Ciências Sociais para estudar padrões de comportamento.

1.1. Motivação

Neste trabalho pretende-se aprofundar conhecimentos acerca das preferências musicais dos utilizadores da rede social Palco Principal¹.

Com base nesta análise é possível obter-se uma caracterização de um conjunto de indivíduos e das suas relações ao longo do tempo. Este género de estudo permite não só o conhecimento inerente a cada indivíduo como à estrutura da rede que o envolve.

¹ Palcoprincipal.sapo.pt

1.2. Objetivos

O objetivo deste trabalho é a caracterização e análise da evolução de uma rede social, aplicando métodos de estudo de Dinâmica de Redes Sociais. Para esse efeito foi disponibilizada uma base de dados relativa às preferências de estilos musicais dos membros da rede social denominada por Palco Principal². Pretende-se analisar como evoluem as preferências dos utilizadores desta rede. Aplicando os métodos desenvolvidos neste trabalho, podemos obter informação útil para, a título de exemplo, identificar possíveis novas tendências musicais que possam surgir.

Inicialmente é feita a caracterização estática da rede através de técnicas de análise multivariada convencionais e através de técnicas alternativas, de onde se salienta o recurso ao método denominado de *Blockmodeling*. A ferramenta computacional utilizada será o *R*, nomeadamente as packages *tnet*, *blockmodeling* e *igraph*.

1.3. Contribuições

Durante a realização deste trabalho foi elaborada uma metodologia baseada na análise de redes sociais, com incidência na deteção de comunidades e na evolução das redes... O estudo das comunidades é feito com recurso a diferentes algoritmos já existentes, bem como à combinação de algoritmos.

Existe ainda o estudo da dinâmica desta rede do ponto de vista das medidas obtidas a nível de rede e indivíduo assim como das comunidades formadas pelos diferentes algoritmos.

O resultado prático deste trabalho evidencia diferenças entre redes binárias e redes ponderadas com o exemplo da rede em questão. É também proposta uma interpretação dos resultados obtidos para valores de várias medidas usadas na análise deste tipo de dados. Por fim é feita uma comparação entre diferentes tipos de algoritmos de deteção de comunidades assim como a análise de evolução dos *clusters* obtidos por um método específico (*Blockmodeling*).

² Disponível em: <http://Palcoprincipal.sapo.pt>

1.4. Organização

Este trabalho está dividido em 5 capítulos. No primeiro é feita uma breve introdução e contextualização da temática inerente a esta dissertação. No segundo capítulo é apresentado o estado da arte respeitante a conceitos sobre redes considerando medidas tanto a nível de Rede como a nível de Vértice para grafos ponderados e não ponderados assim como uma perspetiva do estudo de evolução de redes.

O terceiro capítulo resume alguns métodos de deteção de comunidades e deteção de eventos. O quarto capítulo começa por descrever a metodologia que vai ser aplicada para o estudo da rede em questão. De seguida são apresentados os resultados obtidos pela aplicação dos diferentes algoritmos descritos nos capítulos anteriores bem como a sua discussão. Por fim, o Capítulo 5 apresenta conclusões bem como perspetivas de trabalho futuro.

Capítulo 2

Análise de Redes Sociais

Atualmente o conceito de Rede Social encontra-se intimamente relacionado com o estudo de redes de relações profissionais ou pessoais que foram difundidas com base no surgimento de sistemas como o *Facebook*, *LinkedIn* e *Twitter*. No entanto, este conceito é bem mais antigo que a própria internet.

Numa visão muito simplista, podemos dizer que qualquer grupo de pessoas com relacionamento/ interesses em comum constitui uma Rede Social. Vamos então definir uma Rede Social como um conjunto finito de nós e arestas que representam Atores e Relações existentes entre eles. (Wasserman and Faust, 1994). A Figura 1 apresenta um exemplo simples da representação duma Rede Social.

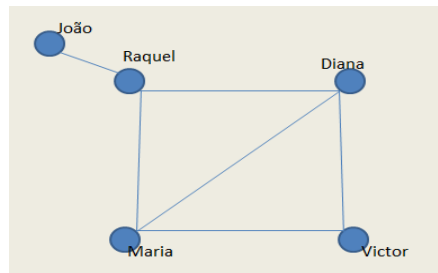


Figura 1 - Grafo simples

Neste exemplo, os atores são pessoas (representadas por círculos), e as relações (representadas por linhas que ligam dois círculos) podem significar, por exemplo, a existência de amizade entre eles. Isto é, se duas pessoas estão ligadas então são amigas.

Na realidade, estas ligações podem representar muitas outras coisas tais como qualquer tipo de relação pessoal desde mero conhecido a familiar ou relação profissional. Pode ainda representar fluxo de informação. Se em vez de pessoas considerarmos que os atores são outro tipo de entidades, continuamos a estar perante uma rede social. O

exemplo mais simples é os atores serem empresas e as relações entre estas serem de, por exemplo clientes, fornecedores, parceiros, etc.

Com o intuito de aprofundar o conhecimento destas ligações e os padrões que lhe estão subjacentes têm-se vindo a desenvolver os mais variados estudos nas últimas décadas. Cada vez mais as Redes Sociais estão a ser usadas em estudos nas mais diversas áreas, desde a Saúde ao Marketing. Nos próximos capítulos vão ser introduzidos conceitos que nos permitem representar e analisar situações reais recorrendo à Teoria de Grafos.

2.1. Teoria de Grafos

De uma forma abstrata, ou seja, independentemente do que está a ser representado, há uma estrutura semelhante a todas as redes. Existe sempre um conjunto de atores e possíveis relações entre eles. Matematicamente, esta estrutura denomina-se Grafo. Neste capítulo vamos ver alguns tipos de grafos e as medidas utilizadas para melhor conhecer/ compreender cada estrutura. A terminologia difere dependendo da área do conhecimento. O contexto de Teoria dos Grafos denomina atores como vértices e as ligações como arestas.

Um grafo G é então definido como um diagrama constituído por um conjunto de vértices, que se podem ligar ou não por arestas. Os vértices extremidade de cada aresta dizem-se adjacentes. Cada aresta une exatamente dois vértices.

De uma maneira mais formal, um grafo G com n vértices e m arestas define-se como um conjunto ordenado $(V(G), E(G), \phi G)$, em que $V(G)=\{v_1, v_2, \dots, v_n\}$ representa o conjunto de vértices (não vazio), $A(G)=\{e_1, e_2, \dots, e_m\}$ é o conjunto de arestas e ϕG faz corresponder a cada aresta os seus extremos $\phi G(e_p) = \{\{v_i, v_j\}, v_i, v_j \in V(G) \text{ e } v_i, v_j \text{ são extremos de } e_p\}$ ((Bondy, 1976) segundo Oliveira e Gama (2011)).

Por questões de simplificação vamos apenas considerar grafos com número de vértices finito.

O estudo mais antigo que se conhece neste tema é de Leonhard Euler em 1736, com a solução do problema das pontes de Königsberg. Uma descrição mais detalhada deste

problema e seu contexto histórico pode ser consultada no livro “The Early Mathematics of Leonhard Euler” (Sandifer, 2007).

Além da representação da Figura 1, pode usar-se também a representação matricial à qual chamamos Matriz de Adjacência. Nesta forma, vamos ter uma matriz quadrada que cruza os vértices consigo próprios. No caso mais simples aparece um 1 se existe uma aresta a ligar os vértices e um 0 se estes não se ligam. Esta representação encontra-se na Figura 2.

	João	Raquel	Diana	Vitor	Maria
João	0	1	0	0	0
Raquel	1	0	1	0	1
Diana	0	1	0	1	1
Vitor	0	0	1	0	1
Maria	0	1	1	1	0

Figura 2 - Matriz de Adjacência do grafo da Figura 1

2.1.1. Conceitos sobre Grafos

Neste capítulo vamos começar por definir alguns conceitos importantes sobre medidas usadas na caracterização de uma rede social (Aldous and Wilson 2000).

- **Grau do vértice** é o número de vértices com ligação a ele, ou seja, um vértice tem grau m se nele incidem m arestas.
- Um **Loop** é uma aresta que liga um vértice a ele próprio.
- Se dois vértices se ligam por mais do que uma aresta, então, são ligados por **arestas múltiplas**.
- Um **Grafo Simples** não apresenta arestas múltiplas nem *loops*.
- Um **Subgrafo** de um Grafo G é um subconjunto de vértices e arestas do Grafo G .
- Um **Passeio** de comprimento k do vértice u para o vértice v é uma sucessão de k arestas de tal forma que o vértice final de uma aresta é o vértice inicial da aresta seguinte. Esta sequência começa no vértice u e acaba no vértice v .
- Um **Caminho** é um passeio em que nenhum vértice nem nenhuma aresta entre o vértice inicial e o final se repetem.
- O **Comprimento** de um caminho é o número de arestas que fazem parte desse caminho.

- O **Caminho mais curto** entre dois vértices, também denominado de **Distância Geodésica**, é o menor caminho entre dois vértices, isto é, o caminho de menor comprimento.
- Um **Ciclo** é um caminho fechado, isto é, o vértice inicial é o vértice final.
- Um Grafo G diz-se **Conexo** se para qualquer par de vértices pertencentes a $V(G)$, existe um caminho que os liga. Caso contrário o Grafo é Desconexo.
- Cada parte conexa de um Grafo **Desconexo** denomina-se Componente.

2.1.2. Grafos direcionados (Digrafos)

Como foi referido antes, os vértices ligam-se através de arestas. Se a cada aresta está associada uma direção ou sentido, então essa ligação designa-se por arco (Nooy, Mrvar e Batagelj, 2005), que pode ser representado por uma seta associada a um sentido. Um Grafo em que existem arcos chama-se Grafo direcionado ou Digrafo. O conjunto de arcos é definido por $A(G) = \{(v_i, v_j), v_i, v_j \in V(G)\}$. Este tipo de Grafos é muito útil para representar situações como a rede de fornecimento entre empresas ou um mapa que mostra o sentido das ruas, como se ilustra na Figura 3.

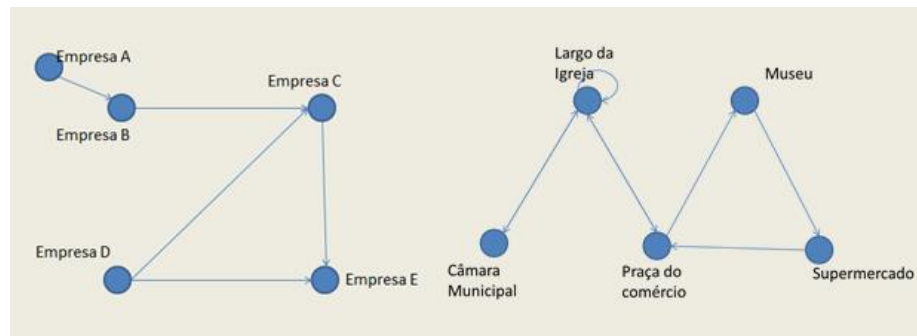


Figura 3 - Rede de fornecimento entre empresas (à esquerda) e Mapa com informação de sentido de ruas numa localidade (à direita)

Na Figura 3 (à direita) reparamos que do vértice “Largo da Igreja” sai uma seta que volta para ele próprio. Em Grafos direcionados há esta possibilidade. Neste exemplo específico, pode estar representada uma rua que dá a volta à Igreja.

Neste caso a Matriz de Adjacência não é necessariamente simétrica porque as relações não são recíprocas.

	CM	LI	PC	Museu	Smercado
CM	0	1	0	0	0
LI	1	1	1	0	0
PC	0	1	0	1	0
Museu	0	0	0	0	1
Smercado	0	0	1	0	0

Figura 4 - Matriz de Adjacência do Mapa da Figura 3 (direita)

Cada entrada não nula na Matriz de Adjacência significa a existência de uma relação direcionada do elemento da linha correspondente para o da coluna. O elemento simétrico não é obrigatoriamente igual.

2.1.3. Grafos Ponderados

Se a cada aresta associarmos um valor real obtemos um Grafo Ponderado. Estes Grafos são extremamente úteis para a representação, por exemplo do quanto uma relação entre dois vértices é forte (se os vértices representarem pessoas, os pesos das arestas podem variar de mais alto para pessoas de família e mais baixo para meros conhecidos), ou distâncias (se os vértices representarem cidades os pesos das arestas podem representar a distância entre elas ou o custo de viagem). Como exemplo consideramos o grafo da Figura 3 (à direita) e adicionamos a cada aresta o tempo em minutos que esta demora a ser percorrida como mostra a Figura 5.

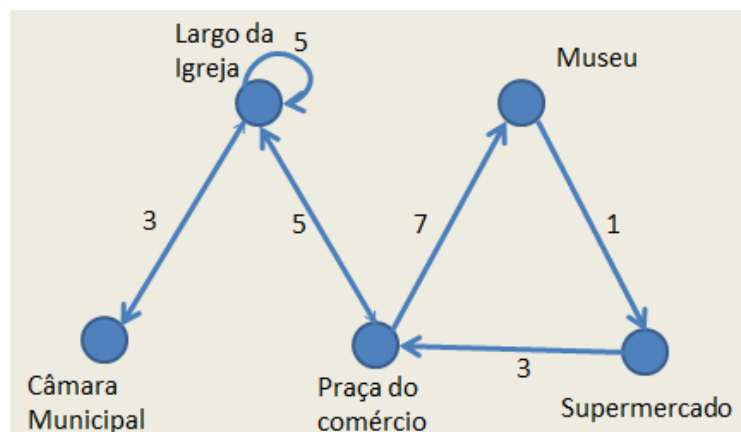


Figura 5 – Exemplo de Grafo Ponderado

Podemos ver, por exemplo, que para chegar da Câmara Municipal ao Museu, demoramos $3+5+7=15$ minutos. A definição de Comprimento do Caminho para estes Grafos não é o número de arestas percorridas mas sim a soma da ponderação destas. A Matriz de Adjacência já não vai ser binária pois a cada entrada não nula vai corresponder o peso da respetiva ligação como podemos observar na Figura 6.

	CM	LI	PC	Museu	Smercado
CM	0	3	0	0	0
LI	3	5	5	0	0
PC	0	5	0	7	0
Museu	0	0	0	0	1
Smercado	0	0	3	0	0

Figura 6 - Matriz de Adjacência do grafo direcionado e ponderado representado na Figura 4

A ponderação das arestas pode também associar maior proximidade a valores mais elevados. No exemplo da Figura 7 que representa o número de horas que em média certas pessoas passam juntas por semana, quanto maior a ponderação da aresta mais próxima está a pessoa.

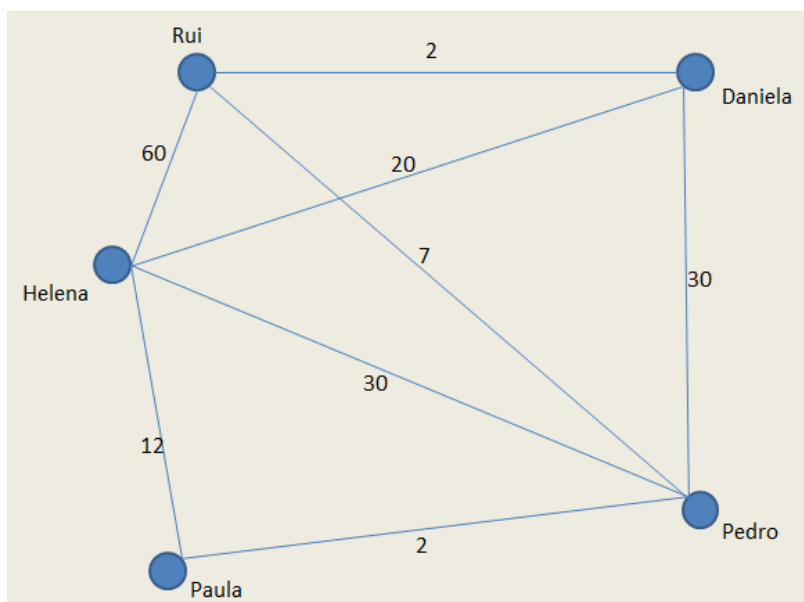


Figura 7 – Grafo ponderado das horas que cinco indivíduos passam juntos por semana

Como poderemos verificar a partir da Secção 2.2, as medidas definidas para redes ponderadas são calculadas a partir de matrizes de distância, como no exemplo da Figura

5 e não de semelhança, como definidas na pelo grafo da Figura 7. Para fazer a transformação de matriz de semelhança para matriz de distância Tore Opsahl³ sugere que seja feita uma normalização dos valores. Esta normalização consiste na divisão dos valores referentes a cada aresta pela média destes valores. De seguida estes valores são invertidos. Desta forma ficam associados valores mais baixos a pessoas mais próximas e valores mais altos às mais distantes. Podemos ver os valores obtidos no grafo da Figura 8.

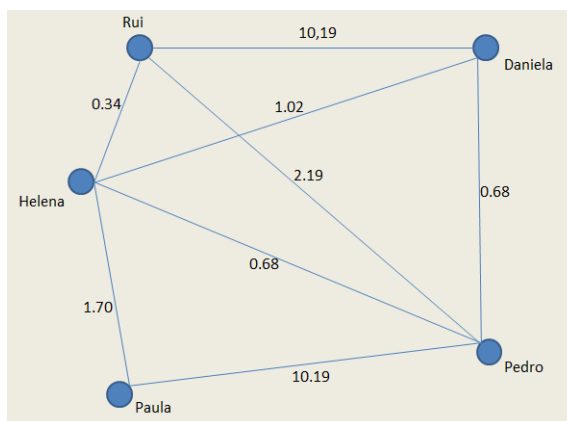


Figura 8 – Grafo ponderado depois da transformação dos valores pelo método sugerido por Tore Opsahl

2.1.4. Grafos completos, nulos e regulares

De acordo com Aldous e Wilson (2000) um grafo diz-se **completo** se existem todas as ligações possíveis entre os seus vértices. Se um grafo com n vértices é não direcionado o número máximo de arestas é dado por:

$$\# A(G) = \frac{n(n - 1)}{2}$$

Se for um grafo direcionado este valor é:

$$\# A(G) = n(n + 1)$$

onde $\# A(G)$ é o número de arestas do grafo G , uma vez que um grafo direcionado pode ter *loops*.

No caso dos grafos **nulos** os seus vértices não apresentam qualquer tipo de relação, não existem portanto arestas. Consequentemente todos os vértices têm grau 0.

³ <http://toreopsahl.com/tnet/weighted-networks/shortest-paths>

No caso dos grafos **regulares**, todos os vértices têm o mesmo grau, ou seja, o mesmo número de vértices vizinhos.

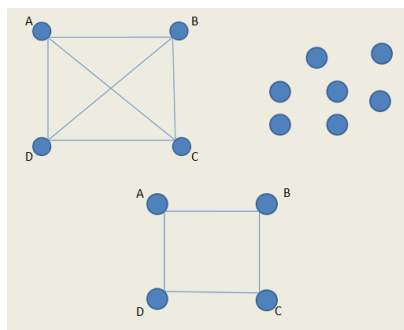


Figura 9 - Exemplo de grafo completo (em cima à esquerda), nulo (em cima à direita) e regular (em baixo).

Na Figura 9, o grafo do canto superior esquerdo além de completo é regular uma vez que todos os vértices apresentam o mesmo grau. Quando um grafo é regular, verifica-se que o resultado da soma de cada linha/coluna da matriz de adjacência é igual. A representação matricial de um grafo completo apresenta o valor 1 em todas as entradas exceto na diagonal principal. Um grafo nulo representa-se por uma matriz nula. O grafo regular da Figura 9 (em baixo) representa-se pela matriz que se mostra na Figura 10.

	A	B	C	D
A	0	1	0	1
B	1	0	1	0
C	0	1	0	1
D	1	0	1	0

Figura 10 - Matriz de Adjacência do Grafo Regular da Figura 9 (em baixo)

2.1.5. Árvores e Grafos Bipartidos

A nível do tipo de estrutura salientamos a estrutura em **Árvore** quando um grafo conexo não apresenta ciclos (Aldous e Wilson 2000). Este tipo de grafos é normalmente usado para a representação de relações de responsabilidade entre vértices. Um exemplo muito conhecido deste tipo de grafos são as árvores genealógicas.

Salientam-se ainda os Grafos Bipartidos cujo conjunto de vértices pode ser dividido em dois grupos de maneira a que todas as arestas liguem vértices de diferentes grupos.

No exemplo da Figura 11 tomamos alguns alunos e professores de uma escola. As arestas correspondem à relação “é aluno de” ou “é professor de”. Do lado esquerdo temos os professores que não se ligam entre si porque não dão aulas uns aos outros. Do lado direito temos o grupo de alunos que também não se ligam entre si pelas mesmas razões. Só existem ligações entre atores dos diferentes grupos. Estas relações são não direcionadas uma vez que são obrigatoriamente recíprocas. (Newman, 2003)

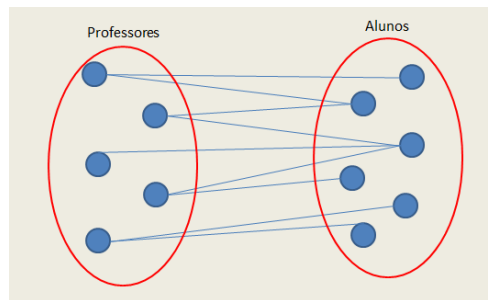


Figura 11 – Grafo Bipartido

2.2. Medidas Estatísticas de Análise de Redes

Na Figura 12 mostramos um grafo constituído por 11 vértices e 15 arestas.

A sequência A-B-C-A-K-J é um passeio. Se tomarmos o mesmo vértice inicial e o mesmo vértice final, um caminho possível é A-B-C-E-H-I-J. O caminho mais curto entre A e J é A-K-J. Este caminho tem comprimento 2 uma vez que percorre 2 arestas.

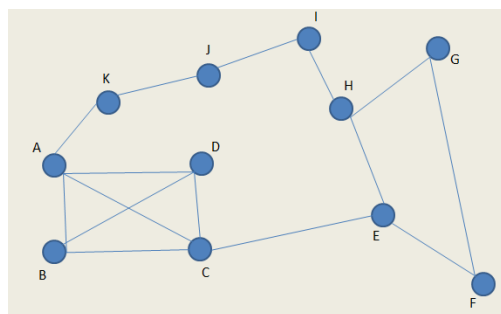


Figura 12 - Exemplo de um grafo

2.2.1. Medidas a nível dos vértices (atores)

2.2.1.1. Excentricidade

Tomemos o conjunto de caminhos mais curtos desde um vértice a todos os outros vértices alcançáveis por este na rede. A Excentricidade do vértice é o comprimento do maior caminho mais curto deste conjunto (Wasserman and Faust, 1994).

Do conjunto de caminhos mais curtos entre o vértice A e todos os outros 10 vértices, (Figura 12), o mais longo é A-C-E-H-G com comprimento 4 (há outro caminho com o mesmo comprimento). Logo o vértice A tem excentricidade 4.

Voltando ao exemplo do grafo ponderado da Figura 7, e tendo presente a noção de caminho mais curto para este tipo de grafos, obtemos a matriz de comprimentos dos caminhos mais curtos da Tabela 1.

	Helena	Rui	Daniela	Pedro	Paula
Helena	0	0.339583	1.01875	0.679167	1.697917
Rui	0.339583	0	1.358333	1.01875	2.0375
Daniela	1.01875	1.358333	0	0.679167	2.716667
Pedro	0.679167	1.01875	0.679167	0	2.377083
Paula	1.697917	2.0375	2.716667	2.377083	0

Tabela 1 - Matriz de distâncias no grafo da Figura 7

A excentricidade de cada vértice é o valor máximo da linha ou coluna que lhe corresponde.

2.2.1.2. Grau

Como foi referido anteriormente, o grau de um vértice é o número de arestas que nele incidem. Na Figura 12 o grau máximo é 4 para os vértices A e C.

Em Grafos direcionados, no caso de existirem loops, estes são contados duas vezes. Podemos ainda definir o *indegree* que contabiliza as relações que apontam para o vértice e *outdegree* que contabiliza as relações para as quais o vértice aponta (Aldous and Wilson 2000).

Para grafos ponderados esta medida tem tido diferentes interpretações. Uma delas é a soma das ponderações de todas as arestas ligadas ao vértice (Newman 2004). Tore

Opsahl (2010) propõe uma medida com um factor de ponderação entre o número de ligações do vértice e o valor associado a cada aresta, que se traduz por:

$$c_D^{w\alpha}(i) = k_i \times \left(\frac{s_i}{k_i}\right)^\alpha = k_i^{(1-\alpha)} \times s_i^\alpha$$

onde k_i é o número de vizinhos do vértice i e s_i é a soma das ponderações das arestas correspondentes e α é um parâmetro que define se é dada mais importância a k_i ou a s_i .

	alpha=0	alpha=0.5	alpha=1	alpha=1.5
Helena	4	22.09072	122	673.767
Rui	3	14.3875	69	330.9124
Daniela	3	12.49	52	216.4933
Pedro	4	16.61325	69	286.5785
Paula	2	5.291503	14	37.04052

Tabela 2 - Grau para diferentes valores de α do exemplo da Figura 7

Observando a Tabela 2 vemos que, para valores de $\alpha = 0$ e $\alpha = 1$ obtemos o número de vizinhos e a soma das ponderações das arestas incidentes em cada vértice, respetivamente. Para $\alpha = 0.5$, é dada maior importância ao número de vizinhos e para $\alpha = 1.5$ é dada maior importância à ponderação das arestas. Nos dois casos, Helena apresenta valores mais altos nesta medida e Paula valores mais baixos. No entanto, é de salientar a troca de posições (2ª e 3ª posição) entre Pedro e Rui em relação aos valores de $\alpha=0$ (Pedro – 4, Rui – 3) e $\alpha=1.5$ (Pedro - 286.58, Rui – 330.91).

2.2.1.3. Intermediação (Betweenness) de um vértice

A intermediação de um vértice é a soma das proporções entre o número de caminhos mais curtos entre dois vértices E e A que passam por um vértice U e o número de caminhos mais curtos existentes entre estes vértices ((Bondy, 1976) segundo Oliveira e Gama (2011)). A medida mostra o quanto este vértice é intermediário entre outros vértices. Quanto mais alto este valor mais importante é o papel deste vértice sob este ponto de vista.

Esta medida é definida por:

$$b_U = \sum_{E,A \in V(G) \setminus U} \frac{\sigma_{EA}(U)}{\sigma_{EA}}$$

onde $\sigma_{EA}(U)$ representa o número de caminhos mais curtos entre o vértice E e o vértice A que passam por U exceto aqueles em que U é vértice final ou inicial, e σ_{EA} representa o número de caminhos mais curtos no Grafo que ligam E a A .

Os vértices que apresentam alta intermediação são normalmente designados de *Gatekeepers*.

A Figura 13 mostra a posição de tais vértices na rede.

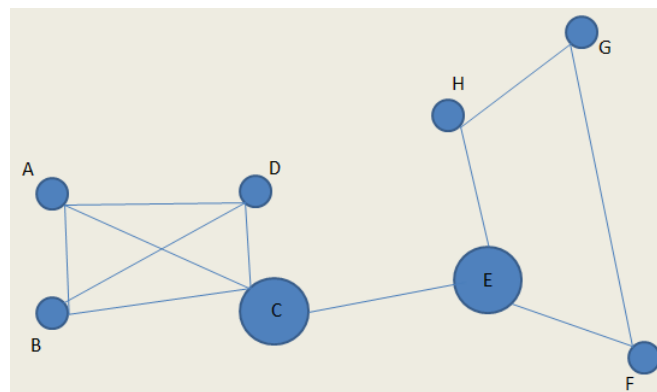


Figura 13 - Exemplo de *Gatekeepers*

Na Figura 13, os vértices C e E são *Gatekeepers* uma vez que a ligação entre eles é a única via de comunicação entre o subgrafo $\{A, B, C, D\}$ e o subgrafo com os restantes vértices.

Podemos estender este conceito às arestas. Definimos desta forma o quanto uma aresta é importante para a intermediação entre os atores desta rede. No mesmo exemplo a ligação entre C e E é chamada de Ponte uma vez que se for retirada o grafo passa a ser desconexo. De uma forma geral, uma Ponte é uma aresta que se for retirada vai fazer com que aumente o número de componentes do Grafo (Harary, Noman, e Cartwright (1965) segundo Granovetter (1973)). Por outras palavras deixa de haver possibilidade de comunicação entre duas ou mais partes de um Grafo.

Uma aresta é uma Ponte Local se, ao ser retirada o grafo não fica desconexo mas os caminhos mais curtos que por ela passavam se tornam mais longos.

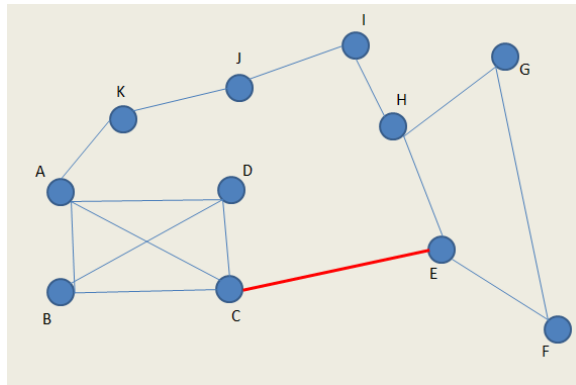


Figura 14 - Exemplo de Ponte Local

Ao retirar a aresta que une os vértices C e E do grafo da Figura 14, os caminhos entre muitos dos vértices vão tornar-se bastante mais longos. Logo esta aresta é uma Ponte Local.

Para grafos ponderados, Tore Opsahl et al, (2010) propõe uma definição à semelhança da definição de Grau, em que há um parâmetro α que regula a importância entre o número de vizinhos e ponderação das arestas.

$$c_B^{w\alpha}(i) = \frac{g_{jk}^{w\alpha}(i)}{g_{jk}^{w\alpha}}$$

onde $g_{jk}^{w\alpha}(i)$ é o número de caminhos mais curtos entre os vértices j e k que passam por i e $g_{jk}^{w\alpha}$ é o número de caminhos mais curtos entre os mesmos dois vértices. A única diferença é definição de caminho mais curto que neste caso tem em conta a ponderação das arestas e o parâmetro α .

2.2.1.4. Proximidade

Esta medida mede a proximidade de um vértice aos restantes vértices que consegue alcançar na rede. Quanto mais alto este valor mais central é a posição ocupada por este vértice e com mais rapidez ele atinge qualquer outro vértice (Wasserman and Faust, 1994).

O seu valor obtém-se por:

$$cl_v = \frac{n - 1}{\sum_{u \in V(G) \setminus v} d(u, v)}$$

onde n é o número de vértices e $d(u, v)$ é a distância geodésica entre u e v .

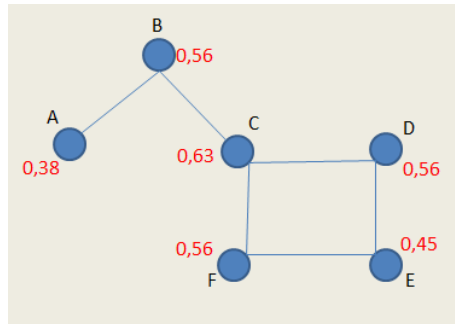


Figura 15- Exemplo dos valores de proximidade entre vértices

No exemplo da Figura 15 o vértice C é o que apresenta uma proximidade superior. Logo este vértice é o mais rápido a alcançar os outros vértices do grafo.

Mais uma vez Tore Opsahl (2010) dá uma definição de Proximidade em redes ponderadas onde a diferença está na forma como são calculadas as distâncias entre vértices.

$$c_C^{w^\alpha}(i) = \left[\sum_{j=1}^N d^{w^\alpha}(i, j) \right]^{-1}$$

Analogamente às definições anteriores aparece o parâmetro α que pondera a importância do número de vizinhos e a ponderação das arestas.

2.2.1.5. Coeficiente de Clustering Local

Para cada vértice medimos a densidade da rede formada pelos seus vizinhos (Watts e Strogatz (1998) segunda Oliveira e Gama (2011)). O seu valor obtém-se por:

$$C_i = \frac{2|e_{vu}|}{k_i(k_i - 1)} : v_u, v_v \in N_i, e_{vu} \in E$$

onde N_i é a vizinhança de v_i (vértices com ligação a v_i), E é o conjunto de arestas em N_i , $|e_{vu}|$ é o número de arestas em E e k_i é o grau de v_i .

Voltando à Figura 12 vamos calcular o Coeficiente de *Clustering* Local para o vértice A.

A vizinhança deste vértice é mostrada na Figura 16.

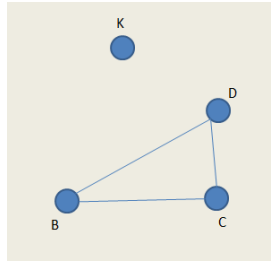


Figura 16- Vizinhança do vértice A

$$Clustering(A) = \frac{2 \times 3}{4 \times 3} = \frac{2}{4} = 0,5$$

Esta medida estima a probabilidade de dois vizinhos de um vértice estarem ligados entre si.

Mais uma vez esta definição é generalizada por Tore Opsahl (2010) da seguinte forma:

$$C_{\omega} = \frac{\text{valor total de tercetos fechados}}{\text{valor total de tercetos}} = \frac{\sum_{\tau \Delta} \omega}{\sum_{\tau} \omega}$$

Tercetos são conjuntos de três vértices ligados por duas (tercetos abertos) ou três (tercetos fechados) arestas. O valor de cada terceto pode ser calculado de quatro formas diferentes, média aritmética (am), média geométrica (gm), máximo (ma) e mínimo (mi) das ponderações das arestas do terceto. Voltando ao exemplo da Figura 7 obtemos, para os valores de *Clustering* Local, os resultados da Tabela 3, com os valores de cada terceto calculados das quatro maneiras referidas acima e ainda ignorando a ponderação das arestas (bi).

	am	gm	ma	mi	bi
Helena	0.7158470	0.7401169	0.6923077	0.7735849	0.6666667
Rui	1	1	1	1	1
Daniela	1	1	1	1	1
Pedro	0.8019324	0.8531301	0.7643312	0.9200000	0.6666667
Paula	1	1	1	1	1

Tabela 3 - Coeficiente de *Clustering* Local para o exemplo da Figura 7

Rui, Daniela e Paula apresentam o valor 1 na Tabela 3 porque todos os seus vizinhos estão ligados entre si.

Estes valores são obtidos tendo em conta a rede que cada vértice forma apenas com os seus vizinhos.

2.2.2. Medidas a nível de Rede

2.2.2.1. Raio e Diâmetro numa rede

Tomemos os valores de Excentricidade de todos os vértices de uma rede.

O Raio dessa rede é o valor mais baixo deste conjunto e o Diâmetro dessa rede é o valor mais alto. Este valor (Diâmetro) corresponde ao comprimento do maior caminho mais curto (Wasserman e Faust, 1994).

Voltando à Figura 12, os maiores caminhos mais curtos de todos os vértices têm comprimento 4. Logo 4 é o Diâmetro e o Raio desta rede.

2.2.2.2. Densidade da Rede

A densidade de uma rede é a razão entre o número de arestas dessa rede e o número máximo de arestas possíveis tendo em conta o número de atores na rede (Nooy et al, 2005).

No grafo da Figura 12 temos 15 arestas. Como existem 11 vértices, o número máximo de arestas neste Grafo seria de 55. A densidade é então de 0,27.

2.2.2.3. Distância Geodésica Média

Distância Geodésica Média é a razão entre a soma dos comprimentos de todos os caminhos mais curtos e o número de arestas possíveis para um grafo completo com o mesmo número de vértices. Esta medida mostra o quanto em média os atores estão separados entre si.

$$AGD = \frac{1}{\frac{1}{2}n(n+1)} \sum_{u \geq v} d_{uv}$$

em que d_{uv} representa a distância entre os vértices u e v (Newman, 2003).

Para o grafo da Figura 17

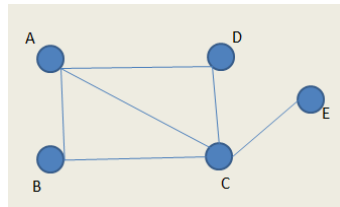


Figura 17- Grafo com 6 vértices

a distância geodésica média é dada por:

$$AGV = \frac{1 + 1 + 1 + 2 + 1 + 2 + 2 + 1 + 1 + 2}{21} = \frac{14}{21} = 0,66$$

2.2.2.4. Grau Médio

Medida que mostra em média qual o grau de cada vértice do grafo. Dá a ideia do quanto os atores estão ligados entre si.

No exemplo da Figura 17 o Grau Médio é dado por

$$GM = \frac{3 + 2 + 4 + 2 + 1}{5} = \frac{12}{5} = 2,4$$

2.2.2.5. Coeficiente de Clustering Global

O Coeficiente de *Clustering* Global obtém-se pela média dos valores dos coeficientes de *Clustering* Local.

Retomando o exemplo das Tabela 3 os valores do Coeficiente de *Clustering* da rede completa estão apresentados na Tabela 4.

am	gm	ma	mi	bi
0.8250905	0.8316951	0.8549223	0.8160377	0.7894737

Tabela 4 - Coeficiente de *Clustering* Global

Note-se a diferença de cerca de 7 centésimas entre as diferentes maneiras de calcular o valor dos tercetos na Tabela 4.

2.3. Estudo da Dinâmica de Redes

O estudo da dinâmica de redes permite incorporar a variável tempo na análise de uma rede. A evolução de uma rede ao longo do tempo pode ser analisada em termos de métricas ao nível do ator e da rede.

Seja D um conjunto de dados dividido em t subconjuntos, de forma que cada subconjunto D_i corresponde a um instante. Porém, analisar a evolução de um conjunto de dados temporais a partir de uma sucessão de momentos temporais pode sujeitar a análise a perturbações casuais, e não garante a estacionariedade dos dados. Para contornar esta dificuldade propõem a agregação dos momentos t em janelas (j). (Gama, 2012)

Cada janela (j) analisada resulta da agregação de observações em instantes temporais (t_i a t_k), correspondendo a um conjunto de dados (W_j) que respeita a esse período temporal e, por sua vez, está na base da construção de uma rede (G_j).

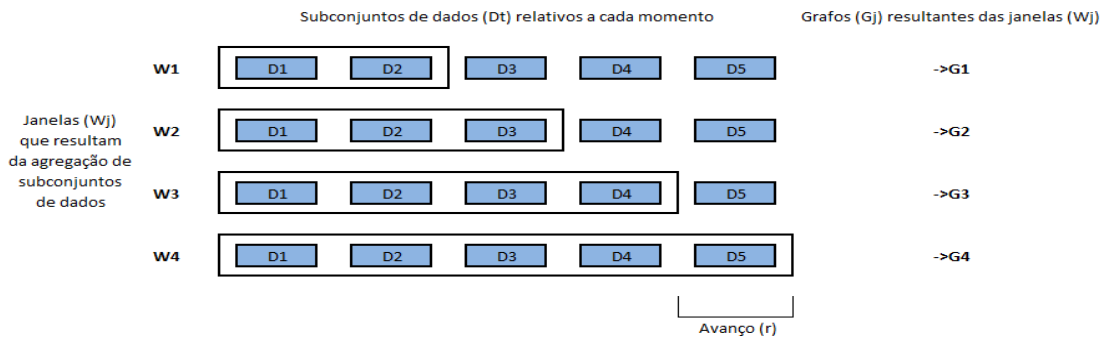
Na literatura podemos encontrar as duas abordagens frequentemente utilizadas para a construção das janelas:

Janelas Acumuladas - O conjunto de dados (D) é subdividido em n subconjuntos (W_j , $j=1, \dots, n$ - que vulgarmente herdam o nome de “janelas”), compreendendo o período temporal decorrido entre o momento 1 ($i=1$) e o momento k_j . Assim, obtemos uma sucessão de “janelas” de dimensão crescente, com um mínimo de inclusão de k_1 momentos e um máximo de t , (sendo t o número total de momentos). O número total de janelas resultantes vai depender do ritmo da sucessão (r), ou seja, o avanço que cada sucessiva “janela” representa em relação à que lhe antecede em termos do último momento considerado em cada uma (assim $k_{j+1}=k_j+r$).

Janelas Deslizantes - O conjunto de dados é dividido em “janelas” (W'_j), de dimensão constante (d). A primeira na sucessão (W'_1) compreende os momentos 1 a d , ou seja $i_1=1$ e $k_1=d$. Sucede-lhe W'_2 , sendo $i_2=i_1+r$ e $k_2=k_1+r$, que mantém a dimensão d da anterior W'_1 . Esta sucessão termina quando $k_n=t$, ou seja, W'_n corresponderá aos momentos de i_n a t .

A Figura 18 mostra um possível esquema para a construção destas janelas.

Janelas Acumuladas



Janelas Deslizantes

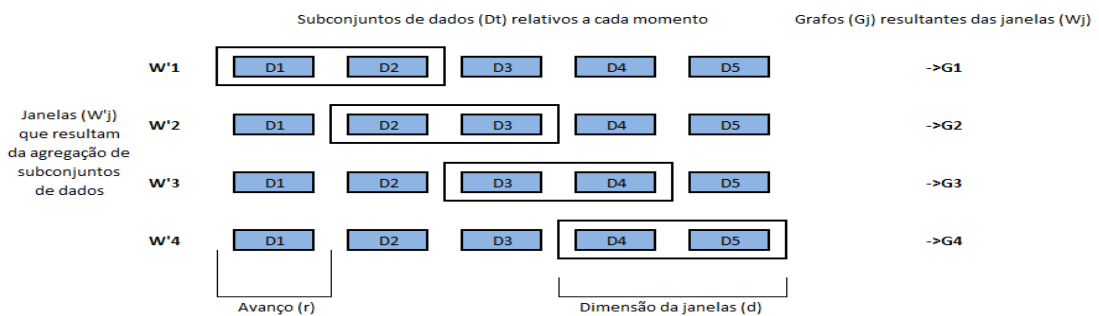


Figura 18 – Janela Acumulada, W_j (em cima) com avanço r e Janela Deslizante, W'_j (em baixo) com avanço r e dimensão d

Capítulo 3

Deteção de Comunidades

A deteção de comunidades é uma outra perspetiva do estudo de redes. A divisão da rede em subgrupos e a análise da evolução temporal destes grupos proporciona conhecimento sobre a rede que não é abrangido pelas técnicas de análises descritas no Capítulo 2. No presente capítulo vamos descrever algumas metodologias que permitem obter este tipo de conhecimento.

Na vida real definimos em geral comunidade como um grupo de indivíduos que apresentam características em comum. Do ponto de vista de redes sociais este conceito é semelhante.

O objetivo da Deteção de Comunidades é dividir o conjunto dos vértices em grupos homogéneos com base nas suas características. Estas características podem ser por exemplo os valores que cada vértice apresenta para as medidas referidas no Capítulo 2. Em geral, dentro duma comunidade existe alta densidade de ligações ao contrário do que acontece entre comunidades distintas onde as ligações são esparsas.

De entre os métodos de Deteção de Comunidades propostos na literatura podemos distinguir diferentes tipos, e.g. : Métodos Tradicionais, Algoritmos Divisivos, Algoritmos baseados na Modularidade, Algoritmos baseados em Inferência Estatística, Algoritmos Espetrais ou Algoritmos Dinâmicos (Fortunato, 2010). Alguns exemplos destes tipos de algoritmos estão enunciados nos pontos seguintes.

3.1. Blockmodeling

A pesquisa de estrutura de uma rede baseia-se nas ligações que esta apresenta. De seguida vamos ver como fazer isso utilizando *Blockmodeling*.

O *Blockmodeling* é um método baseado em inferência estatística cujo objetivo é organizar um conjunto de vértices de uma rede numa partição, que representa um conjunto de classes. Cada conjunto de vértices é habitualmente designado por *cluster* (Nooy, et al, 2005). Cada um dos vértices (atores) pertencentes ao mesmo *cluster*, apresenta ligações idênticas. Isto é, ligam-se aos mesmos vértices ou a vértices idênticos. Esta divisão da rede é feita tendo em conta as ligações entre atores e não as características de cada um. É então um trabalho de pesquisa de estrutura de rede. Como veremos mais adiante, a representação matricial é mais intuitiva para perceber exatamente qual o procedimento. Vamos começar por definir alguns conceitos.

No exemplo das Figuras 19 e 20 temos uma partição em *clusters* de uma Rede com 12 vértices e 22 ligações. A matriz de adjacência (Figura 20) é simétrica, uma vez que as ligações são não direcionadas.

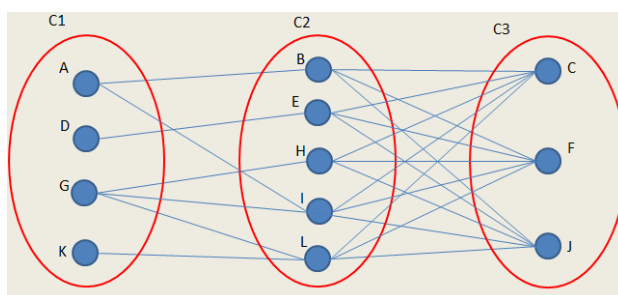


Figura 19 – Partição de um grafo em 3 *clusters*

	A	D	G	K	B	E	H	I	L	C	F	J
A	0	0	0	0	1	0	0	1	0	0	0	0
D	0	0	0	0	0	1	0	0	0	0	0	0
G	0	0	0	0	0	0	1	1	1	0	0	0
K	0	0	0	0	0	0	0	0	1	0	0	0
B	1	0	0	0	0	0	0	0	0	1	1	1
E	0	1	0	0	0	0	0	0	0	1	1	1
H	0	0	1	0	0	0	0	0	0	1	1	1
I	1	0	1	0	0	0	0	0	0	1	1	1
L	0	0	1	1	0	0	0	0	0	1	1	1
C	0	0	0	0	1	1	1	1	1	0	0	0
F	0	0	0	0	1	1	1	1	1	0	0	0
J	0	0	0	0	1	1	1	1	1	0	0	0

Figura 20- Matriz de Adjacência do Grafo da Figura 19

Veremos de seguida alguns conceitos importantes de *Blockmodelling*, tais como os Tipo de Relação e Tipos de Equivalência que podemos procurar na estrutura da rede.

3.1.1. Bloco

Um Bloco contém as células da Matriz de Adjacência que pertencem ao cruzamento de um consigo próprio ou ao cruzamento de dois *Clusters* (Nooy, Mvar e Batagelj, 2005). No exemplo da Figura 20, estão evidenciados os blocos formados. Neste exemplo podemos observar blocos **nulos** quando os seus elementos são todos 0, **completos** quando todos os elementos são 1 ou **regulares** quando existe pelo menos um 1 em cada linha e em cada coluna. Estes tipos de blocos, denominados Blocos Ideais podem ser observados na Figura 20. Note-se que esta matriz está dividida em 9 blocos, 5 blocos nulos, 2 blocos completos e 2 blocos regulares. Existem ainda outros tipos de blocos que não vão ser considerados no presente trabalho.

3.1.2. Tipos de relações

Podemos definir muitos tipos de relações entre *clusters*, mas para o âmbito deste trabalho vamos apenas definir os conceitos de relação regular, completa e nula, segundo Doreian, et al. (2005).

3.1.2.1. Relação Completa

Dois *clusters* X e Y apresentam uma Relação Completa entre si se para cada vértice do *cluster* X existe uma ligação a todos os vértices do *cluster* Y. Na Figura 19 verificamos que existe este tipo de ligação entre os Grupos C2 e C3. A matriz de adjacência entre estes dois *clusters* é uma matriz de 1's (Figura 20).

De uma maneira formal, dizemos que se X e Y são dois *clusters* e R é o conjunto de relações entre os vértices de cada *cluster*, e o tipo de relação define-se por Completa, se:

$$com(X, Y: R) \equiv \forall x \in X \forall y \in Y: (x \neq y \Rightarrow xRy)$$

onde $com(X, Y: R)$ representa uma relação completa e xRy indica a existência de relação entre os atores x e y .

3.1.2.2. Relação Regular

Se a Relação é Regular, então, todos os atores de um *cluster* X , estão ligados a pelo menos um ator do *cluster* Y e vice-versa. No exemplo da Figura 19 podemos verificar este tipo de relação entre os *clusters* $C1$ e $C2$. Se atentarmos na parte da matriz de adjacência (Figura 20) que mostra estas ligações, vemos que existe pelo menos um 1 em cada linha e em cada coluna. Na verdade esta é a condição necessária e suficiente para haver uma ligação de tipo regular entre dois grupos.

Formalmente temos que:

$$reg(X, Y: R) \equiv (\forall x \in X \exists y \in Y: xRy) \wedge (\forall y \in Y \exists x \in X: yRx)$$

onde X e Y são dois *clusters*, $reg(X, Y: R)$ representa uma relação regular e xRy indica a existência de relação entre os atores x e y .

3.1.2.3. Relação Nula

A Relação é Nula se não existem ligações entre os vértices dos diferentes *clusters*. Na Figura 19 os *clusters* $C1$ e $C3$ apresentam esta relação. A matriz de adjacência entre os dois grupos é uma matriz de zeros como podemos ver na Figura 20. Formalmente temos que:

$$nul(X, Y: R) \equiv R(X, Y) = \emptyset$$

onde X e Y são dois *clusters*, $nul(X, Y: R)$ representa uma relação nula e xRy indica a existência de relação entre os atores x e y .

3.1.3. Tipos de Equivalência

Ao associar alguns Tipos de Relação definimos Tipos de Equivalência.

Se na procura de uma partição de uma rede apenas permitirmos Blocos Nulos ou Completos, então estamos à procura de um *blockmodel* que apresente **Equivalência Estrutural**.

Quando na procura da partição são permitidos Blocos Nulos e Regulares então procuramos uma **Equivalência Regular**.

Note-se que Equivalência Estrutural pode ser considerado um caso particular de Equivalência Regular uma vez que Relação Completa é um caso particular de Relação Regular.

3.1.4. Blockmodel

Um *Blockmodel* atribui uma classe a cada vértice de uma rede e especifica os tipos de relações permitidas dentro e entre classes (Nooy, et al., 2005).

O objetivo do *Blockmodeling* é descobrir os padrões de estrutura de grafos e fazer a divisão do conjunto de vértices em classes. O trabalho é realizado sobre a Matriz de Adjacência. Por isso este método não é aconselhado para redes com grandes dimensões. Muito raramente uma rede apresenta características que encaixam de forma perfeita nos conceitos de equivalência apresentadas. Daqui surgiu a necessidade de criar métodos capazes de quantificar o quanto uma solução é melhor do que outra. Para tal é feita a contagem do número de células da matriz que não combinam com o que seria de esperar num bloco perfeito. Este resultado designa-se erro.

Para melhor descrever o algoritmo e o conceito de erro atentemos na Figura 21.

	A	B	C	D	E	F	G	H	I	J	K	L
A	0	1	0	0	0	0	0	0	1	0	0	0
B	1	0	0	0	0	1	0	0	0	1	0	0
C	0	0	0	0	1	0	0	1	1	0	0	1
D	0	0	0	0	1	0	0	0	0	1	0	0
E	0	0	1	1	0	1	0	0	0	1	0	0
F	0	1	0	0	1	0	0	1	1	0	0	1
G	0	0	0	0	0	0	0	1	1	0	0	1
H	0	0	1	0	0	1	1	0	0	1	0	0
I	1	0	1	0	0	1	1	0	0	1	0	0
J	0	1	0	1	1	0	0	1	1	0	0	1
K	0	0	0	0	0	0	0	0	0	0	0	1
L	0	0	1	0	0	1	1	0	0	1	1	0

Figura 21 - Matriz de Adjacência de uma Rede

Numa primeira fase deve haver um estudo da rede pois só desta forma podemos saber qual o número de *clusters* em que faz sentido dividir o conjunto de atores e quais os tipos de estrutura permitidos. O processo pode começar por agrupar de forma aleatória o

conjunto de vértices no número de subgrupos apropriados ou fazer a otimização de uma partição inicial gerada por outro método de *Clustering*. Depois vão ser efetuadas uma série de permutações (trocas de linhas/colunas da Matriz de Adjacência). Em cada permutação o erro é calculado contando quantas ligações deviam existir e não existem ou não deviam existir mas existem para que a matriz do modelo encaixasse perfeitamente na matriz constituída por Blocos Ideais tendo em conta as estruturas permitidas. O processo pára quando de uma permutação para a outra o erro não baixa ou quando é atingido o número máximo de iterações permitido.

O resultado da aplicação deste processo na matriz da Figura 21 está representado na Figura 22. Aqui verificamos que a relação entre os vértices J e D, e a falta de relação entre os vértices C e B impedem que os Blocos em que estão inseridos sejam estruturalmente equivalentes.

	A	D	G	K	B	E	H	I	L	C	F	J
A	0	0	0	0	1	0	0	1	0	0	0	0
D	0	0	0	0	0	1	0	0	0	0	0	1
G	0	0	0	0	0	0	1	1	1	0	0	0
K	0	0	0	0	0	0	0	0	1	0	0	0
B	1	0	0	0	0	0	0	0	0	0	1	1
E	0	1	0	0	0	0	0	0	0	1	1	1
H	0	0	1	0	0	0	0	0	0	1	1	1
I	1	0	1	0	0	0	0	0	0	1	1	1
L	0	0	1	1	0	0	0	0	0	1	1	1
C	0	0	0	0	0	1	1	1	1	0	0	0
F	0	0	0	0	1	1	1	1	1	0	0	0
J	0	1	0	0	1	1	1	1	1	0	0	0

Figura 22- Matriz de Adjacência equivalente após permutações

Num processo de otimização, o *blockmodeling* vai, permutação a permutação, tentando melhorar este erro. Este processo não garante que se encontre a solução ótima.

Aceitando a matriz da Figura 22 como solução, podemos então condensar a informação nela contida. Para tal representamos cada Bloco pelo nome do tipo de relação como mostra a Figura 23:

null	regular	null
regular	null	complete
null	complete	null

Figura 23- Matriz reduzida da matriz da Figura 19

Podem ser atribuídos valores às ligações entre *clusters*. Estes valores são estabelecidos tendo em conta o tipo de ligação entre eles e ainda o peso de cada aresta para o caso de grafos ponderados. Ao conjunto de regras que estabelece tais valores Batagelj (1997) chamou *averaging rules*.

3.1.5. Definição Formal

Batagelj (1997) define *Blockmodel* de um grafo G constituído por um conjunto de vértices V e um conjunto de arestas E , como um vector $\mathcal{M} = (U, K, T, Q, \pi, \mu)$ onde:

- U é um conjunto de tipos de Unidades (tipo de vértices)
- $K \subseteq U \times U$ é um conjunto de ligações entre tipos de vértices
- T é um conjunto de relações entre *clusters*
- $\pi : K \rightarrow T$ é uma função que a cada par de tipos de unidades associa as relações permitidas
- Q é o conjunto de *averaging rules* (regras mediadoras)
- $\alpha : K \rightarrow Q$ é a função que associa regras de cálculo para os valores da ligação entre dois *clusters*;
- $\mu : V \rightarrow U$ é a função que associa a cada vértice de V o seu tipo.

Então, para $t \in U$:

$$C(t) = \mu^{-1}(t) = \{x \in V : \mu(x) = t\}$$

define um *cluster* com os vértices associados ao tipo t .

Desta forma,

$$\zeta(\mu) = \{C(t) : t \in U\}$$

define uma partição do conjunto de atores V .

Uma função $\mu : V \rightarrow U$ determina um *blockmodel* se e só se

$$\forall (t, w) \in K: \pi(t, w)(C(t), C(w))$$

ou seja, são respeitadas as relações definidas entre os pares de *clusters* e

$$\forall (t, w) \in U \times U \setminus K: nul(C(t), C(w))$$

ou seja, *clusters* que não têm relação definida entre si apresentam relação nula.

A Figura 24 representa o diagrama reduzido do exemplo da Figura 19. Por diagrama reduzido entenda-se diagrama que representa apenas as classes e as suas relações.

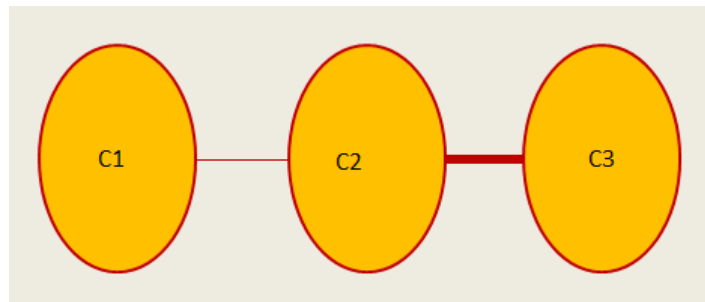


Figura 24- Diagrama reduzido do grafo da Figura 19

Como já foi dito anteriormente, entre as classes C1 e C2 existe Relação Regular que aqui é representada por um traço fino, enquanto, entre C2 e C3, existe uma Relação Completa, representada aqui por um traço mais grosso. Entre os *clusters* C1 e C3 não há ligações (Relação Nula).

Neste exemplo, se considerarmos que o conjunto de vértices contem clientes, supermercados e fornecedores, respetivamente C1, C2 e C3, tem-se $U = \{\text{clientes, supermercados, fornecedores}\}$.

O conjunto K de ligações é $K = \{(C1, C2), (C2, C3)\}$ e os tipos de ligações do conjunto T são $T = \{\text{nula, completa, regular}\}$.

O conjunto Q pode ser definido de várias maneiras, para simplificar vamos apenas definir $Q = \{0, \text{se relação nula}; 1, \text{se relação regular}; 2 \text{ se relação completa}\}$

Então temos $\mathcal{M} = (U, K, T, Q, \pi, \mu)$ definido.

Podemos verificar que \mathcal{M} define um *Blokmodel* no exemplo, uma vez que para cada relação em K entre dois *clusters* existe uma correspondência a um elemento em T ((C1,C2) – Relação Regular e (C2,C3) Relação completa) e a única ligação possível entre os *clusters* que não existe no diagrama (C1,C3) apresenta Relação Nula.

3.1.6. Generalização de Blockmodeling a Redes Ponderadas

Para a aplicação do *Blockmodeling* a Redes Ponderadas, Ziberna (2007) distingue *Valued Blockmodeling* e *Homogeneity Blockmodeling*. Para esta generalização o conceito de blocos ideais tem que ser redefinido. O método *Valued Blockmodeling* baseia-se na substituição de valores por 0 ou 1 dependendo de um parâmetro m . Este parâmetro deve ser escolhido de forma a ser o valor mínimo que caracteriza a relação entre um vértice e um *cluster* ou entre dois vértices, dependendo do tipo de blocos permitido, de forma a que esta ligação satisfaça a condição do bloco. Para o presente trabalho vamos definir apenas o *Homogeneity Blockmodeling*. Como o nome sugere, esta abordagem procura blocos homogêneos. Assim, a definição de bloco nulo em redes binárias mantém-se, ou seja, os elementos neste tipo de bloco devem ser todos nulos. No entanto um bloco completo é definido como um bloco com os valores todos iguais. Também se definem blocos regulares e dominantes que estão no entanto fora do âmbito deste trabalho.

Consideram-se duas formas para o cálculo do erro associado a cada partição, a soma dos quadrados dos desvios à média e a soma dos valores absolutos dos desvios à mediana. Para cada uma destas definições, o cálculo é feito de forma distinta para os diferentes tipos de blocos como podemos verificar na Tabela 5.

Bloco ideal	Soma de quadrados	Desvios Absolutos	Posição do Bloco
Nulos	$\sum_{i,j} b_{ij}^2$	$\sum_{i,j} b_{ij} $	Fora da diagonal
	$\sum_{i \neq j} b_{ij}^2 + SS(diag(B))$	$\sum_{i \neq j} b_{ij} + ad(diag(B))$	Diagonal
Completo	$ss(b_{ij})$	$ad(b_{ij})$	Fora da Diagonal
	$ss(b_{ij}) + SS(diag(B))$	$ad(b_{ij}) + ad(diag(B))$	Diagonal

Tabela 5 – Inconsistências de Bloco em *Homogeneity Blockmodeling*

Na Tabela 5 B refere-se a um Bloco; b_{ij} um elemento da matriz B correspondente ao cruzamento da i -ésima linha com a j -ésima coluna; ss – Soma dos quadrados dos

desvios à média, ad - soma dos desvios absolutos à mediana e $\text{diag}(\mathbf{B})$ – vetor com os elementos da diagonal da matriz \mathbf{B} .

3.2. *Clustering Hierárquico*

O agrupamento de vértices pode também ser feito recorrendo ao *Clustering Hierárquico* (Doreian, et al., 2005). Este é um método tradicional e pode inclusivamente ser usado para análise preliminar à aplicação do *Blockmodeling*.

O primeiro passo consiste em considerar que cada indivíduo forma um *cluster* e construir a matriz de dissemelhança entre os indivíduos (*clusters*). Existem várias possibilidades para o cálculo desta matriz em particular a Distância Euclideana, definida por:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

e a distância de Manhattan definida por:

$$d(x, y) = \sum_{i=1}^n |x_i - y_i|$$

onde x e y , descritos como $x = (x_1, x_2, \dots, x_n)$ e $y = (y_1, y_2, \dots, y_n)$ são dois indivíduos e x_i é o valor da variável i para o sujeito x e y_i é o valor da variável i para o sujeito y .

No segundo passo efetua-se uma procura na matriz de dissemelhanças pelos pares de indivíduos mais próximos para que estes sejam agregados. A matriz de dissemelhança tem que ser atualizada uma vez que houve formação de um *cluster* com dois indivíduos. Esta atualização pode ser feita por diferentes critérios dependendo do problema em questão. Alguns desses critérios são:

- **Menor Distância**, em que a distância do novo *cluster* a um indivíduo é distância mais curta entre os indivíduos do *cluster* e esse indivíduo.

- **Maior Distância**, em que a distância do novo *cluster* a um indivíduo é a maior distância entre os indivíduos do *cluster* e esse indivíduo
- **Distância Média**, em que a distância do novo *cluster* a um indivíduo é a média das distâncias dos indivíduos do *cluster* a esse indivíduo
- **Método do Centróide**, em que as coordenadas do novo *cluster* são as médias das coordenadas dos indivíduos que fazem parte desse *cluster*; as novas distâncias são recalculadas a partir destas coordenadas.

O segundo passo é repetido até que todos os indivíduos pertençam a um mesmo *cluster*. Não existe consenso em relação ao melhor método para recalculas as distâncias. É aconselhável experimentar mais do que um.

De cada vez que se agregam *clusters* é guardada a distância que os separava.

O resultado deste método aglomerativo é em geral visualizado em forma de dendrograma, como mostra a Figura 25, que representa as possíveis divisões da base de dados referente a estatísticas de detenções por agressão, assassinato ou violação em cada um dos 50 estados americanos em 1973.

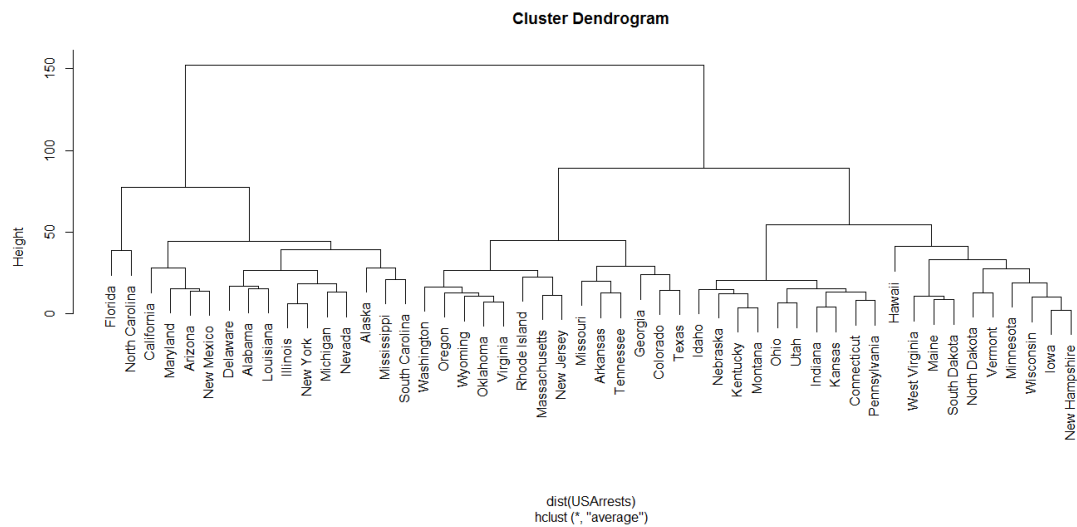


Figura 25 - Exemplo de Dendrograma da aplicação de *Clustering* Hierárquico usando a Distância Euclidiana e o Critério do Valor Médio (USArrests - R database)

Existem também diferentes critérios sobre quantos *clusters* reter. Em geral, por observação do Dendrograma verifica-se qual a maior distância entre os *clusters*. O

exemplo da Figura 25 sugere que se devem reter dois *clusters* (sendo um deles formado pelos primeiros 16 indivíduos do lado esquerdo e o outro pelos indivíduos restantes) uma vez que a altura a que estes se reúnem evidencia uma maior distância entre estes do que entre os restantes.

Existem no entanto outros índices que podem ser calculados para avaliar e decidir a solução a reter.

3.3. Modularidade

O conceito de Modularidade que foi introduzido originalmente por Newman e Girvan (2004) é uma medida da rede respeitante a uma dada partição que mede o quanto essa divisão em comunidades é ou não significativa, isto é, o quanto os *clusters* desta partição são distintos uns dos outros. Quanto mais próximo de 1 melhor é a divisão. Em geral, valores acima de 0.3 são indicadores de estrutura em comunidades. A definição formal é:

$$Q = \frac{1}{2m} \sum_{vw} \left[A_{vw} - \frac{k_v k_w}{2m} \right] \sum_i \delta(c_v, i) \delta(c_w, i)$$

onde m é o número de arestas na rede, v e w são vértices, A_{vw} toma valor 1 se os vértices v e w estão ligados ou 0 caso contrário, k_v é o grau de v e $\delta(c_v, i)\delta(c_w, i)$ toma valor 1 se os vértices v e w estão no mesmo *cluster* (i) e 0 caso contrário.

De uma forma mais simples, esta função compara a densidade de um *cluster* existente na rede objeto de análise, com a densidade esperada numa rede aleatória com o mesmo número de vértices, em que a probabilidade de existir ligação entre dois vértices v e w é $\frac{k_v k_w}{2m}$, o que respeita o grau de cada vértice.

Aquando da generalização para redes ponderadas Newman (2004) define as seguintes modificações: A_{vw} é a ponderação da aresta que liga v a w , k_w é a soma das

ponderações das ligações de w , por fim $m = \frac{1}{2} \sum_{vw} A_{vw}$ representa o valor de metade da soma das ponderações de todas as arestas da rede.

Certos algoritmos de detecção de comunidades baseiam o processo de otimização de partição no aumento da Modularidade.

3.3.1. Algoritmo CNM

Em 2004 Clauset, Newman e Moore propuseram um algoritmo de detecção de comunidades direcionado a redes com número muito elevado de vértices. A ideia, como noutros métodos, é dividir a rede em grupos onde existe uma grande densidade intra-grupo e baixa densidade entre-grupos. Este método usa um processo hierárquico de otimização local da Modularidade. Inicialmente, cada vértice é um *cluster*, a cada passo juntam-se as duas comunidades que provocam o maior aumento (ou menor diminuição) na Modularidade. O processo pára quando não for possível continuar a aumentar a Modularidade.

3.3.2. Walktrap

Este é um método de Detecção de Comunidades do tipo dinâmico que usa o conceito de **passeios aleatórios**. Um passeio aleatório acontece quando, desde o vértice inicial, o vértice seguinte é escolhido aleatória e uniformemente entre os seus vizinhos (Pons e Latapt, 2005). A sequência de vértices que define um passeio aleatório é uma Cadeia de Markov uma vez que a escolha do vértice seguinte não é influenciada pela escolha dos vértices anteriores. Se uma rede está bem definida em comunidades, um passeio aleatório vai tender a ficar dentro de uma comunidade uma vez que existem menos arestas que o levam para fora dessa comunidade do que arestas que estão dentro. Desta forma são efetuados vários passeios aleatórios pelo grafo e guardada informação que vai possibilitar o cálculo da probabilidade de um passeio entre cada par de vértices da rede ter um certo comprimento fixo. A partir desta informação é calculada uma matriz de distâncias entre todos os pares de vértices que associa valores mais baixos a vértices no mesmo *cluster* e valores mais altos a vértices em *clusters* diferentes.

Depois deste processo é aplicado o *Clustering* Hierárquico como descrito no ponto 3.2.

Também neste método a significância da partição encontrada é baseada na Modularidade (Oliveira e Gama, 2011).

3.3.3. Algoritmo de Girvan and Newman

Um exemplo de algoritmo divisivo é o algoritmo de Girvan e Newman (2004). Este algoritmo usa o conceito de intermediação de arestas (*edge betweenness*). É também um método hierárquico que vai decompondo o grafo à medida que vai retirando as arestas por ordem decrescente do valor de intermediação de cada uma. Como foi visto no ponto 2.2.1.3. as arestas que apresentam valores mais altos nesta medida são muito provavelmente pontes ou pontes locais. Ao retirar gradualmente estas arestas o grafo vai sendo subdividido em comunidades. Por cada passo os valores de intermediação das arestas são recalculados. Este algoritmo pára quando já não existem mais arestas para retirar. A Modularidade pode também aqui servir de critério de escolha da melhor partição.

3.4. Evolução de Redes - Detecção de Eventos

Uma outra forma de estudo da dinâmica de uma rede é a análise do que acontece com as comunidades dessa rede ao longo do tempo. Esta forma de análise permite observar as mudanças na estrutura da rede. Formação e Dissolução de comunidades são só dois exemplos do que pode acontecer. Desde a Biologia à Sociologia estes fenómenos podem ser evidenciados de uma forma muito visual por este tipo de análise.

O estudo é efetuado a partir de informação da rede respeitante a certos instantes de tempo devidamente ordenados. Cada *snapshot* pode representar apenas o que aconteceu desde o instante da *snapshot* anterior ou pode conter a informação acumulada das *snapshots* desde o ponto inicial da análise.

3.4.1. Algoritmo de Choobdar, et al

Choobdar, et al (2012) propõe um método faseado em dois passos. O primeiro é a divisão da rede em comunidades por aplicação de um método de detecção de comunidades apropriado a uma matriz cujos valores se referem medidas a nível de vértice descritas no Capítulo 2. Por fim, a cada nó é associado um tipo, por exemplo, o número do *cluster* no qual ficou inserido. O segundo passo é a definição e identificação de eventos relacionados com o tamanho das comunidades que ocorrem ao longo do tempo. Entenda-se por tamanho da comunidade o número de indivíduos inseridos nessa comunidade. Estes autores definem cinco tipos de eventos que podem acontecer a um grupo de um instante para o instante seguinte:

- **Crescimento**, quando um grupo mostra uma tendência constante de aumento de tamanho num intervalo de tempo.
- **Diminuição**, quando um grupo mostra uma tendência constante de diminuição de tamanho num intervalo de tempo.
- **Formação**, quando o grupo não existia no instante de tempo anterior.
- **Dissolução**, quando um grupo que existia no instante anterior não existe no instante actual.
- **Continuação**, quando um *cluster* não altera o seu tamanho de um instante para o outro; embora possa observar-se aumento ou diminuição do número de vértices, o tamanho não se altera significativamente.

Todos os tipos de eventos que ocorrem a cada instante são detetados e descritos. Depois é elaborado um conjunto de regras gerado de forma a descrever cada evento. Este é um processo com duas fases. Na primeira fase são detetados todos os intervalos de tempo em que ocorrem eventos, tais intervalos são denominados de intervalos de transição.

Para detetar intervalos de transição, isto é, intervalos de tempo em que as alterações de tamanho de um *cluster* C_i sejam significativas, os autores propõe o seguinte processo: em primeiro lugar o número de vértices de C_i ao longo do tempo é guardado na forma de série temporal $F_i(t)$ com $t \in [1, T]$, sendo T o número de *snapshots* disponíveis. Os intervalos de transição são extraídos por segmentação desta série temporal quando comparada à reta obtida por regressão linear entre o tamanho dos *clusters* e o tempo. Se

num intervalo $[a,b]$ o erro da reta ajustada à série, medido em termos da soma do quadrado dos resíduos, ultrapassar um certo limite, o intervalo é encurtado para o ponto j , sendo $j < b$, o ponto onde é dada a melhor aproximação para $F_i(a, j - 1)$. Este processo pode ser controlado pelo número máximo de intervalos de transição. Para tal, o valor do erro máximo aumenta o seu valor até que o número de intervalos não seja superior ao pretendido. Neste processo também pode ser definido o erro máximo, não havendo assim restrição para o número de intervalos de transição.

A inclinação dos segmentos em cada intervalo revela o comportamento do *cluster*.

Na segunda fase, vão-se registar as regras de transição correspondentes a cada intervalo de transição. Estas regras são do tipo $C_i \rightarrow C_j$ e mostram que nós do grupo C_i passam para o grupo C_j no intervalo correspondente. A contagem suporte de uma regra deste tipo refere-se ao número absoluto de vértices que apresentaram esta movimentação neste intervalo.

3.4.2. MONIC

Spiliopoulou, et al. (2006) propõe um modelo um pouco diferente do identificado anteriormente. Este modelo não é específico para redes, e necessita que os *clusters* já estejam formados.

Este método incorpora uma função de envelhecimento que associa um valor entre 0 e 1 a cada vértice em cada instante de tempo. Desta forma um valor mais baixo pode ser associado a dados mais antigos para que estes não tenham tanta influência em resultados mais atuais. A partir desta função é calculado o valor de *overlap* entre *clusters* definido por

$$overlap(X, Y) = \frac{\sum_{a \in X \cap Y} age(a, t_j)}{\sum_{x \in X} age(x, t_j)}$$

onde t_i e t_j representam instantes de tempo e $i < j$, a e x representam vértices, X é um *cluster* no instante t_i e Y é um *cluster* no instante t_j e $age(a, t_j)$ é o valor da função de envelhecimento do vértice a no instante t_j .

Neste caso, as transições que ocorrem podem ser externas ou internas ao *cluster*. Ao nível externo, e dependendo do valor de *overlap*, o *cluster* pode:

- **Sobreviver** caso $overlap(X, Y)$ seja máximo para qualquer *cluster* no instante t_j e seja superior a um valor limite (τ) estabelecido para este efeito.
- **Dividir-se** se existem diferentes *clusters* (y_u) no instante t_j que apresentam valor de $overlap(X, Y_u) > \tau_{split}$, inferior a τ .
- **Ser absorvido** quando dois *clusters* distintos no instante t_i apresentam o valor de *overlap* superior a τ significa que um deles foi absorvido.
- **Desaparecer** caso nenhuma das situações anteriores se verifique.
- **Nascer** quando não são resultado de nenhuma transição externa.

Algumas destas definições vão ao encontro das apresentadas no método anterior. Já a nível interno um *cluster* que sobrevive de um instante para o instante seguinte pode: mudar de tamanho (encolher ou expandir-se), tornar-se mais ou menos compacto, mudar de localização ou não apresentar qualquer mudança. De referir que apenas as transições internas vão ser exploradas no presente trabalho.

Capítulo 4

Caso de estudo: Palco Principal

Neste capítulo pretende-se aplicar os conhecimentos teóricos referidos ao longo deste trabalho a um caso prático da rede social Palco Principal³. Inicialmente vão ser calculadas medidas a nível de rede e de vértice a duas matrizes referentes ao primeiro e quinto instantes de tempo de que dispomos e a matrizes que representam as Janelas Deslizante e Acumulada. A partir destes resultados será feita uma análise da evolução da rede. De seguida vão ser aplicados métodos de deteção de comunidades às mesmas matrizes (Janelas Deslizante e Acumulada) e será feito o estudo da evolução de comunidades.

4.1. O que é o Palco Principal

Desde 2009 em Portugal, o Palco Principal⁴ é uma rede social voltada para música com origem em Portugal e que neste momento está também no Brasil, Moçambique, Angola e Cabo Verde.

Existem três tipos de utilizadores do Palco Principal: os ouvintes, os artistas e entidades ligadas à música tais como produtores de espetáculo, editores, rádios ou lojas de instrumentos.

Todos os membros do “Palco Principal” têm ao dispor as novidades do mundo da música. Há uma actualização permanente de toda a informação acerca de eventos musicais tais como concertos, lançamentos musicais, festivais ou passatempos. Os membros têm a possibilidade de fazer *download* de milhares de músicas dos mais variados estilos e criar a sua própria *playlist*. Podem ainda deixar a sua opinião acerca

⁴ palcoprincipal.sapo.pt

dos trabalhos dos artistas. A sua rede de contactos pode ser aumentada tanto a nível nacional como internacional facilitando assim o fluxo de informação.

Ao registar-se como ouvinte, o utilizador cria uma página pessoal onde partilha informação, fotos, vídeos ou preferências musicais. Ao ligar-se a outros utilizadores pode trocar opiniões, conhecer novas bandas nos seus estilos preferidos e novos estilos.

Todos os músicos têm aqui uma oportunidade de divulgar o seu trabalho, pondo disponível para *download* a sua música ou divulgando as datas das suas atuações. Ao dar a conhecer o seu trabalho pode verificar a sua popularidade e as opiniões dos fãs. Ao interagir com outros membros da rede, trava conhecimento com outros artistas na área. Esta troca de contactos é muito vantajosa na medida em que facilita a procura de artistas com características específicas para um projecto.

As entidades têm aqui uma oportunidade de publicitar o seu trabalho e verificar tendências, ganhar novos clientes e aumentar a sua lista de contactos.

4.2. Informação disponível para o estudo

O Palco Principal é uma rede social onde cada indivíduo pode dar a conhecer as suas preferências em estilos musicais. Na Figura 26 podemos observar o *interface* onde são escolhidas estas preferências.

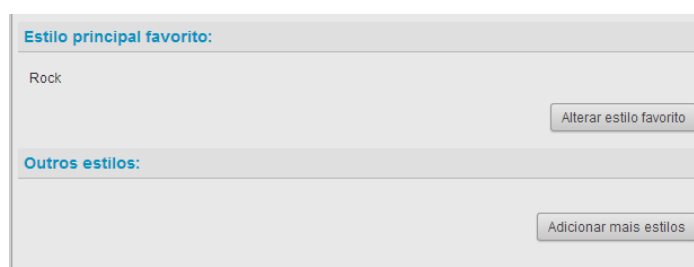


Figura 26 – Parte do *Interface* do site “Palco Principal” referente ao registo de preferências dos utilizadores (palcoprincipal.sapo.pt)

Neste caso, o estilo principal preferido é “Rock”. Neste campo apenas é permitida a escolha de um estilo. Na janela seguinte (“Outros estilos”) podem ser escolhidos tantos estilos quantos desejados. Ambos os campos podem ficar em branco, não dando desta forma o utilizador a conhecer as suas preferências.

Os dados disponíveis para o presente trabalho são cinco matrizes referentes a observações mensais da rede, correspondentes aos períodos de setembro, outubro, novembro e dezembro de 2011 e janeiro de 2012. Estas matrizes são o resultado da agregação de matrizes de adjacência de redes binárias e bipartidas que ligam os utilizadores da rede aos estilos de música. A agregação destas matrizes resultou em 5 matrizes de cruzamento entre estilos de música que a cada entrada associa o número de pessoas que nas suas preferências apresenta os estilos do respetivo cruzamento. A informação contida em cada matriz está acumulada desde de que se iniciou a rede. As redes correspondentes a estas matrizes são redes ponderadas e conexas uma vez que ligam todos os nós. Cada matriz apresenta 45 linhas e 45 colunas e representa a Matriz de Adjacência de uma rede com 45 vértices. Destes 45 vértices, 44 representam diferentes estilos de música e 1, identificado como “null”, representa a ausência de preferência musical. Nas células da diagonal de cada matriz encontra-se o número de pessoas que escolheu como estilo principal preferido o estilo de música referente a esse cruzamento. Na Figura 27 apresenta-se parte da matriz do primeiro momento.

O1	Rock	Tunas	Outro	Rap/Hip-Hop	R&B	Metal
Rock	8982	91	2902	7036	1225	2642
Tunas	91	279	56	181	19	14
Outro	2902	56	1928	14223	2751	439
Rap/Hip-Hop	7036	181	14223	60451	13935	1179
R&B	1225	19	2751	13935	2520	269
Metal	2.642	14	439	1.179	269	3.336

Figura 27 – Parte da matriz de preferências referentes a setembro de 2011

Na primeira entrada da matriz verificamos que existem 8982 utilizadores cuja preferência principal é “Rock”. A soma da diagonal principal dá-nos o número de utilizadores da rede. As outras células, que não estão na diagonal principal, referem-se às preferências alternativas ou sub-estilos. Cada utilizador pode escolher tantas quantas

entender. Cada par de estilos vai contribuir com peso 1 para as entradas da matriz onde se cruzam os estilos. Note-se que o mesmo utilizador pode contribuir para mais do que uma entrada da matriz. Por exemplo, um utilizador que apresente nas suas preferências “Rock”, “Tunas” e “Metal” contribui com peso 1 para cada um dos pares “Rock”-“Tunas”, “Rock”- “Metal” e “Tunas”- “Metal”. As matrizes são portanto simétricas. Estes dados podem ser representados em forma de rede social. Cada estilo de música é um vértice e o valor de cada aresta representa o número de utilizadores que nas suas preferências apresenta o par formado pelos dois vértices do extremo da respetiva aresta. Os valores associados aos *loops* representam o número de pessoas cujo estilo musical preferido é representado pelo vértice que apresenta o *loop*. Na Figura 28 podemos ver o aspeto da Rede no primeiro momento depois de retirados os *loops*. O tamanho dos vértices é proporcional ao respetivo grau.

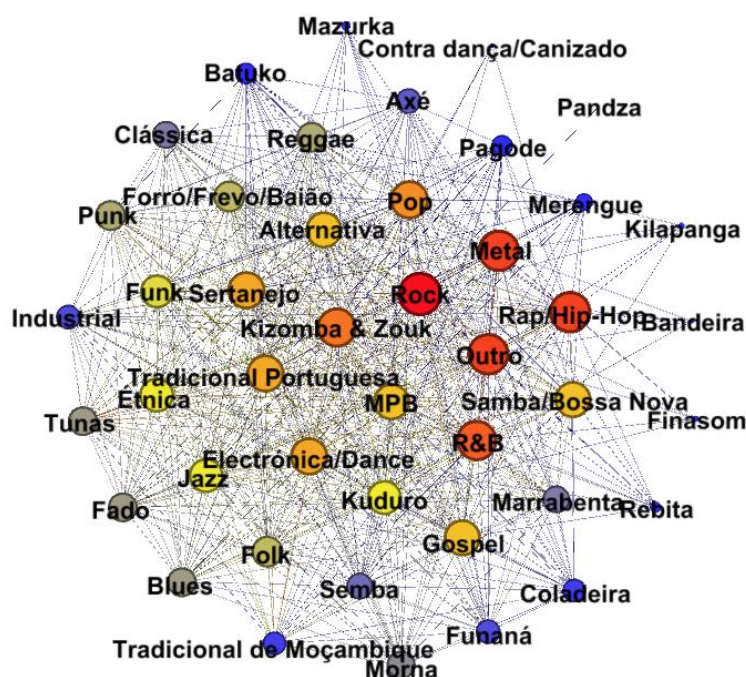


Figura 28 - Diagrama da rede referente a setembro de 2011 de onde retiraram os *loops*

Antes de proceder à análise de resultados vão especificar-se quais as atitudes dos utilizadores que podem influenciar as mudanças na rede:

- Entrada de novos utilizadores na rede que escolhem preferências musicais. Por cada par de estilos escolhido por um utilizador a ligação correspondente entre todos os utilizadores aumenta (ou é criada) com peso 1.
- Mudança nas preferências dos utilizadores já na rede. Tenhamos presente que o acréscimo de um estilo musical nas preferências de um utilizador vai aumentar com peso 1 a ponderação de cada aresta que une o vértice que o representa e os vértices representantes dos estilos que já estavam escolhidos como preferidos por esse utilizador. Se não existir aresta entre estes dois vértices será criada uma nova com peso 1.
- De forma análoga, se um utilizador retira estilos das suas preferências retira peso às arestas correspondentes sendo que, se a aresta tivesse peso 1, o utilizador em questão era o único responsável pela sua existência e esta deixa de existir.

4.3. Metodologia aplicada na elaboração deste estudo

Nesta secção vamos descrever a metodologia de estudo utilizada.

Os dados disponíveis encontram-se organizados em cinco matrizes que vamos designar por O_1, O_2, O_3, O_4, O_5 , referentes às preferências musicais de utilizadores da Rede Social do *site* do Palco Principal em Setembro, Outubro, Novembro e Dezembro de 2011 e Janeiro de 2012.

Na primeira fase deste estudo vamos calcular e comparar medidas a nível de rede e de vértice para as matrizes O_1 e O_5 . Além de detetar possíveis mudanças na rede entre estes períodos temporais, esta análise tem também o intuito de evidenciar as diferenças que as ponderações das arestas provocam nestas métricas. Para esta parte do estudo vão ser aplicadas as definições propostas por Tore Opsahl uma vez que O_1 e O_5 são matrizes de semelhança referentes a redes ponderadas. A *package tnet* do R revela-se central nesta análise uma vez que tem em conta a ponderação das arestas para o cálculo de certas medidas usadas neste estudo.

Na perspectiva de aprofundar a análise da dinâmica desta rede vão ser extraídas matrizes, a partir de O_1, O_2, O_3, O_4 e O_5 , representantes de Janelas Deslizantes, $D_i, i = 1, \dots, n$ e Janelas Acumuladas, $A_i, i = 1, \dots, n$, sendo n o número de janelas de cada tipo. Às matrizes, D_i , e A_i , vai ser aplicada a metodologia descrita para as matrizes O_1 e O_5 . Por esta análise pretende-se perceber, por um lado, se está a haver mudanças nas preferências de estilos de música dos ouvintes (janela deslizante), por outro lado, se essas mudanças, caso existam, são suficientes para mudar a rede como um todo (janela acumulada).

Numa segunda fase a rede vai ser estudada sob o ponto de vista de comunidades. Vão ser aplicados algoritmos de Deteção de Comunidades como o algoritmo CNM sendo a partição resultante destes métodos a partição inicial usada para aplicação de otimização pelo algoritmo *Blockmodeling*, a D_i e A_i . Pretende-se perceber as diferenças e complementaridades entre os métodos assim como as suas possíveis limitações. Para tal, vai ser criada uma partição inicial, P_1 aplicando o algoritmo de CNM. Esta partição vai servir de partição inicial para uma nova partição por aplicação da otimização por *Blockmodeling* (PB_1) A Figura 29 esquematiza este processo.

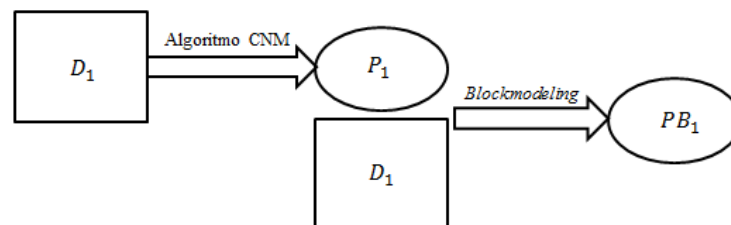


Figura 29 – Esquema de processo de partição pelo algoritmo CNM e posterior otimização pelo algoritmo de *Blockmodeling*

Este processo é repetido para cada janela. Às partições resultantes da otimização por *Blockmodeling* vai ser aplicado o método de deteção de eventos.

Ainda no ponto de vista de comunidades mas numa abordagem um pouco diferente vai ser criada uma partição inicial, C_1 , por *Clustering* Hierárquico aplicado a D_1 . A otimização por *Blockmodeling* vai ser aplicada a D_1 tendo C_1 como partição inicial. A

partição resultante deste processo, CB_1 , vai ser a partição inicial para uma nova otimização por Blockmodeling mas agora aplicada a D_2 . Este processo é repetido em todas as janelas como esquematiza a Figura 30.

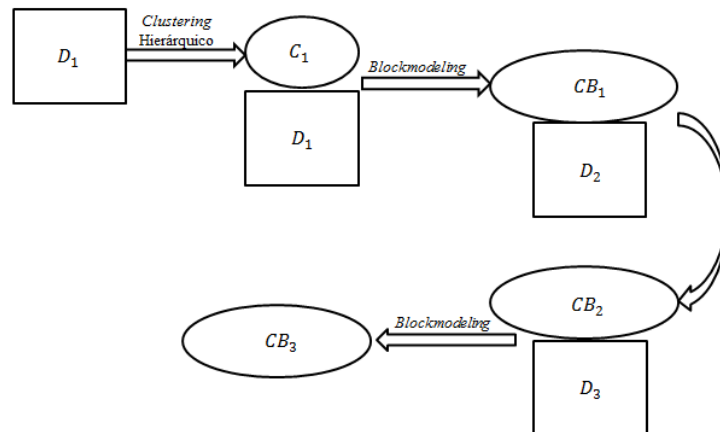


Figura 30 - Esquema do processo de partição por *Clustering Hierárquico* e posterior otimização por *Blockmodeling*

Para a deteção de eventos vamos usar um método baseado nos métodos descritos no Capítulo 3. Este método usa as definições de transições externas do método *MONIC*, no entanto não usa a função *overlap*. Esta foi substituída pela percentagem de vértices nas interseções de *clusters* das diferentes janelas D_1 , D_2 e D_3 .

Para definir que um *cluster sobrevive*, deve existir um *cluster* na janela seguinte com pelo menos 60% dos vértices deste *cluster* e estes devem constituir pelo menos 50% dos vértices deste *cluster*. Um *cluster divide-se* se não existir nenhum *cluster* na janela seguinte com pelo menos 60% dos seus vértices existindo dois *clusters* com pelo menos 40% destes vértices. Um *cluster é absorvido* quando existem dois *clusters* de uma janela que apresentam no mínimo 60% dos seus vértices no mesmo *cluster* da janela seguinte. O que apresentar menor número de vértices na interceção com vértices do *cluster* da janela seguinte foi absorvido. Caso um *cluster* não tenha sofrido nenhuma das

transições acima referidas então o *cluster desapareceu*. Os *clusters* que **nascem** são aqueles que não são resultado de nenhuma transição.

4.4. Extração de dados a partir da informação disponível

Como foi referido na Secção 4.2 foram-nos disponibilizados dados referentes às preferências musicais dos utilizadores da rede social Palco Principal. Cada uma das 5 matrizes, O_1, O_2, O_3, O_4 e O_5 , refere-se a diferentes momentos e a informação armazenada refere-se à acumulação de informação desde o início da rede. Quer isto dizer que cada uma das matrizes pode ser vista como uma janela acumulada. Para retirar mais proveito das técnicas que pretendemos aplicar neste estudo, especialmente para a análise da Janela Deslizante aplicaram-se algumas operações sobre estas matrizes. Na Figura 31 podemos observar o esquema destas operações. Por subtração de cada matriz pela matriz correspondente à informação do instante exatamente anterior foram extraídas as matrizes de transição entre instantes. Este processo dá origem a quatro matrizes de transição, T_1, T_2, T_3, T_4 . Estas matrizes armazenam informação das mudanças entre os meses de setembro e outubro de 2011 outubro e novembro de 2011, novembro e dezembro de 2011 e dezembro de 2011 e janeiro de 2012 respetivamente.

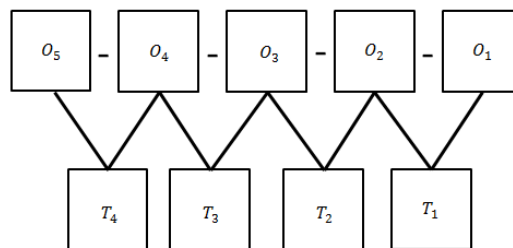


Figura 31 - Operações efetuadas nas matrizes O_1, O_2, O_3, O_4, O_5

Para obter as Janelas Deslizante e Acumulada retiram-se as perdas de ligação (valores negativos das matrizes de transição) uma vez que, pela sua quase inexistência foram considerados irrelevantes para o estudo.

As Janelas Deslizantes foram obtidas somando as matrizes de transição duas a duas em instantes consecutivos como mostra a Figura 32.

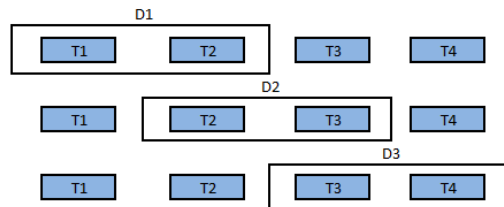


Figura 32 - Construção das Janelas Deslizantes

As matrizes D_1 , D_2 e D_3 referem-se respectivamente ao aumento das ponderações ou criação de arestas entres os meses setembro a novembro de 2011, outubro a dezembro de 2011 e novembro de 2011 a janeiro de 2012 respectivamente.

Para a obtenção das Janelas Acumuladas começamos com a soma das matrizes T_1 e T_2 à semelhança de D_1 . A segunda Janela Acumulada, A_2 , corresponde à soma de A_1 e T_3 . A terceira e última Janela Acumulada, A_3 corresponde à soma de A_2 e T_4 . A Figura 33 ilustra este esquema.

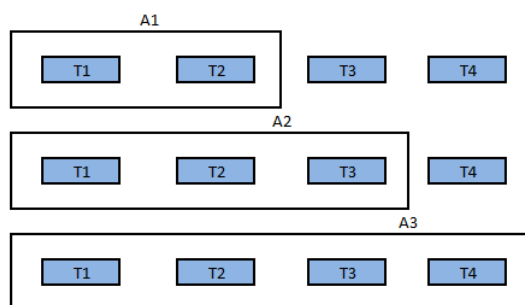


Figura 33 - Construção das Janelas Acumuladas

As janelas A_1 , A_2 e A_3 correspondem às mudanças na rede nos meses setembro, outubro e novembro de 2011; setembro, outubro, novembro e dezembro de 2011; setembro, outubro, novembro, dezembro de 2011 e janeiro de 2012 respectivamente.

Desta forma foram obtidas 3 Janelas Deslizantes e 3 Janelas Acumuladas sendo que a primeira janela em ambos os métodos coincide.

4.5. Estudo das matrizes originais nos primeiro e quinto instantes

Os pontos seguintes descrevem as medidas a nível de Rede e de Vértice para os dados referentes aos primeiro e quinto instantes. Para isso foram usadas as matrizes originais de onde se retiraram os *loops*. Além da comparação de valores nos diferentes instantes são também comparadas as diferenças provocadas pelas ponderações das arestas em relação à rede binária.

4.5.1. Medidas a nível de Rede:

A maneira como os dados são apresentados permite logo uma análise da evolução utilizando a janela acumulada. O instante de tempo 2 (T2) apresenta o que já aconteceu no instante 1 (T1) mais as ligações criadas neste intervalo de tempo. Assim acontece para os tempos seguintes. Os resultados desta análise encontram-se na Tabela 6.

	T1	T2	T3	T4	T5
Nº Utilizadores	92970	97707	100748	104024	106249
Nº Arestas	640	653	657	670	672
Média	226.8017	239.8511	249.1768	263.0315	269.1994
Dist. Média	34.0764	25.5216	26.2964	24.7205	25.2719
Densidade	0.639535	0.653277	0.657505	0.671247	0.673362
Transitividade gm	0.974281	0.977374	0.977533	0.979159	0.98002
Transitividade bi	0.824162	0.833325	0.832503	0.836849	0.839959

Tabela 6 - Medidas de rede ao longo dos 5 Instantes

Podemos verificar que há um aumento do número de utilizadores da rede. Há uma diferença de 13279 utilizadores do 1º para o 5º instante. O aumento mais significativo acontece do instante 1 para o instante 2 e representa cerca de 35% do aumento total. O

aumento menor dá-se do tempo 4 para o tempo 5 e é de cerca de 17% em relação ao aumento total.

O aumento apresentado não significa que não haja pessoas a desistir da rede social. Mostra apenas que há mais pessoas a aderir do que a desistir.

A média das ponderações das arestas está a aumentar ao longo do tempo, este aumento é de cerca de 40 utilizadores para cada ligação. Uma vez que o número de arestas aumenta torna-se evidente que as ponderações também aumentam.

A diminuição da distância média é natural dado o aumento do número de ligações e suas ponderações.

A densidade da rede aumenta o que também é de esperar uma vez que o número de arestas aumenta e o número de vértices é constante.

Para o cálculo do Coeficiente de *Clustering* (Transitividade) mostram-se os resultados obtidos considerando as matrizes binárias (bi) e no caso de considerar a Distância Geométrica (gm). Este método de cálculo da distância foi escolhido uma vez que será tão mais fácil fechar um terceto (conjunto de três vértices e duas arestas) quanto mais pessoas gostem em simultâneo dos tipos de música representados por estes vértices, o que se traduz num maior peso das arestas em questão. Podemos verificar que há uma diferença considerável entre os resultados das duas formas de cálculo. Os valores de Transitividade bi indicam que esta rede apresenta uma grande tendência para o fecho de tercetos, ou seja, criação de novas arestas. Os valores de Transitividade gm reforçam esta tendência uma vez que são superiores e são calculados com base nas ponderações das arestas o que os torna à partida resultados mais realistas.

4.5.2. Medidas a nível dos vértices:

- **Coeficiente de *Clustering* Local**

As medidas a nível dos vértices também são diferentes consoante se dá mais ou menos peso à ponderação das arestas. No caso do Coeficiente de *Clustering* Local podemos ver, nas Tabela 7 e 8, a variação entre o primeiro e último tempo do qual temos

informação. Na Tabela 7 são desprezadas as ponderações das arestas e na Tabela 8 é usada a Distância Geométrica.

Clu_bi	T1	Clu_bi	T5
Rock	0,62237	Rock	0,657807
Rap/Hip-Hop	0,67561	Metal	0,686411
Outro	0,681707	Rap/Hip-Hop	0,712195
Metal	0,681707	Outro	0,718293
R&B	0,710256	Kizomba & Zouk	0,739744

Tabela 7 – Coeficiente de *Clustering* Local ignorando as ponderações das arestas

Os valores estão por ordem crescente. Na Tabela 8 podemos observar os cinco valores mais baixos para o Coeficiente de *Clustering* Local considerando que a rede é binária, isto é, as arestas não são ponderadas. Há um aumento desta medida ao longo do tempo. Podemos explicar este aumento pelo aumento de ligações da rede de vizinhos dos vértices em questão uma vez que a tendência é para aparecimento de ligações e não para o seu desaparecimento. No entanto, o desaparecimento duma ligação pode ter este efeito no caso de no extremo desta ligação haver um vértice que não tem ligação a mais nenhum na rede de vizinhos. Aqui nada pode ser concluído quanto ao peso das arestas.

Clu_gm	T1	Clu_gm	T5
Samba/Bossa Nova	0,93787	Samba/Bossa Nova	0,951526
Metal	0,960244	Metal	0,95898
Rock	0,96247	Rock	0,969872
Tradicional de Moçambique	0,964035	Morna	0,970217
Kizomba & Zouk	0,965763	Tradicional de Moçambique	0,97174

Tabela 8 – Coeficiente de *Clustering* Local calculado pela Distância Geométrica

Na Tabela 8, em que os valores também estão por ordem crescente, podemos verificar que os valores da medida são bastante mais altos do que os da Tabela 7. As posições de uma tabela para a outra também não são as mesmas. Enquanto na Tabela 7 “Rock” é o vértice que apresenta o Coeficiente de *Clustering* mais baixo, na Tabela 8 já é “Samba/Bossa Nova”. Este resultado mostra que, por exemplo, há uma maior tendência para se criarem ligações na rede de vizinhos do vértice “Rock” do que se supunha pelos resultados da Tabela 7.

- **Grau:**

A situação em relação à medida do grau é análoga à anterior. Se não forem consideradas as ponderações das arestas ($\alpha=0$) os cinco vértices com grau mais alto estão na Tabela 9, no entanto, dando mais importância à ponderação da aresta do que ao número de vizinhos do vértice ($\alpha=1.5$), Tabela 10, verificamos que os vértices que apresentam os valores mais altos diferem. O vértice “Metal” deixa de estar nos cinco primeiros (passa mesmo para nono). O estilo “Alternativa”, que no primeiro tempo apresenta 36 vizinhos e no quinto tempo 37 passa a fazer parte dos primeiros 5.

alpha = 0	T1	Grau 0	T5
Rock	43	Rock	43
Outro	41	Metal	42
Rap/Hip-Hop	41	Outro	41
Metal	41	Rap/Hip-Hop	41
R&B	40	R&B	40

Tabela 9 - Grau para rede binária

alpha = 1.5	T1	Grau 1.5	T5
Rap/Hip-Hop	2697721	Rap/Hip-Hop	3515868
Outro	869090,1	Outro	1203619
Rock	756860,3	Rock	993969,9
R&B	584023,8	R&B	770652,9
Alternativa	537119,8	Alternativa	733271,2

Tabela 10 – Grau para rede ponderada

- **Intermediação e Proximidade**

Nas medidas Intermediação e Proximidade há uma análise idêntica. Na Tabela 11 podemos ver quais os cinco vértices que apresentam valores mais elevados para cada situação.

Prox 0	T1	Prox 0	T5
Rock	0,023255814	Rock	0,023255814
Outro	0,022222222	Metal	0,022272723
Rap/Hip-Hop	0,022222222	Outro	0,022222222
Metal	0,022222222	Rap/Hip-Hop	0,022222222
R&B	0,02173913	R&B	0,02173913

Prox 1.5	T1	Prox 1.5	T5
Rap/Hip-Hop	0,0000350	Rap/Hip-Hop	0,0000675
Outro	0,0000350	Outro	0,0000675
R&B	0,0000350	R&B	0,0000675
Rock	0,0000350	Rock	0,0000675
Electrónica/Dance	0,0000350	Electrónica/Dance	0,0000675

Interm 0	T1	Interm 0	T5
Rock	82,82995	Rock	47,70743
Rap/Hip-Hop	36,33828	Metal	36,54076
Outro	30,41328	Rap/Hip-Hop	28,44394
Metal	30,41328	Outro	22,54076
R&B	23,91328	Kizomba & Zouk	22,31104

Interm 1.5	T1	Interm 1.5	T5
Rap/Hip-Hop	803	Rap/Hip-Hop	817
Rock	371,5	Rock	371
Alternativa	42	Metal	42
Kizomba & Zouk	42	Alternativa	42
Trad. Portuguesa	42	Trad. Portuguesa	42

Tabela 11 – Cinco vértices que apresentam valores mais altos de Proximidade com $\alpha=0$ (tabela em cima à esquerda) e $\alpha=1.5$ (em baixo à esquerda) e de Intermediação com $\alpha=0$ (tabela em cima à direita) e $\alpha=1.5$ (em baixo à direita) para os instantes T1 e T5

No estudo dos resultados da Tabela 11 podemos verificar que os vértices mais centrais na rede, vértices que apresentam valores de Proximidade mais elevados são “Rock”, “Rap/Hip-Hop”, “Metal”, “R&B” e “Outro” se considerarmos a rede binária (Tabela 11 à esquerda em cima). Apesar de uma mudança de ordem entre os instantes 1 e 5 os vértices mantêm-se os mais centrais. Já quando dada mais importância à ponderação das arestas (Tabela 11 à esquerda em baixo) o estilo “Metal” deixa de estar nas 5 posições mais altas dando lugar ao estilo “Eletrónica/ Dance”. Aqui esta medida apresenta o mesmo valor para estes 5 vértices. Estes são os vértices que mais rapidamente atingem qualquer vértice na rede. São ideais, por exemplo para difundir algum tipo de publicidade que se destine a toda a rede.

Analisando os valores de Intermediação (Tabela 11 à direita) verificamos que a maioria dos vértices mais centrais na rede são também os vértices que apresentam valores mais altos nesta medida. Isto significa que estes vértices são à partida *Gatekeepers*. A ausência destes provocaria um grande aumento das Distâncias Geodésicas entre os vértices podendo mesmo a rede passar a ser desconexa. Realça-se a diferença entre os resultados para a rede binária (Tabela 11 à direita em cima) e para a rede ponderada quando é dada mais importância às ponderações (Tabela 11 à direita em baixo). Na realidade, o vértice “Rap/Hip-Hop” é mais importante neste sentido do que o vértice “Rock” já que apresenta um valor mais alto nesta medida do que este vértice. Se a matriz fosse binária esta situação era inversa.

4.6. Janela Deslizante e Janela Acumulada para aparecimento/aumento de ligações

No intuito de aprofundar a análise da Dinâmica de Rede aplicou-se análise análoga à da Secção 4.5 às Janelas Deslizante (D_1 , D_2 e D_3) e Acumulada (A_1 , A_2 e A_3) retiradas das matrizes originalmente fornecidas. Como referido na Secção 4.4. este método resultou em 3 Janelas Deslizantes e 3 Acumuladas sendo que as primeiras janelas de cada tipo coincidem.

4.6.1. Janelas Deslizante

Como referido na Secção 2.3 os resultados obtidos com o método Janela Deslizante podem ser influenciados pelos dados mais recentes, uma vez que as observações mais antigas deixam de ser usados nos cálculos.

As Figura 34, 35 e 36 mostram a representação em redes das três Janelas Deslizantes D_1 , D_2 e D_3 . Verificamos que em D_1 existem dois vértices isolados que representam os estilos “Contra dança/Canizado” e “Pandza”. Nas segunda e terceira janelas (D_2 e D_3) existe apenas um vértice que se encontra nessa situação. Em ambos os casos este vértice refere-se ao estilo “Contra dança/Canizado”. Para o cálculo das medidas tanto a nível de rede como de vértice apenas é utilizada a parte conexas do grafo.

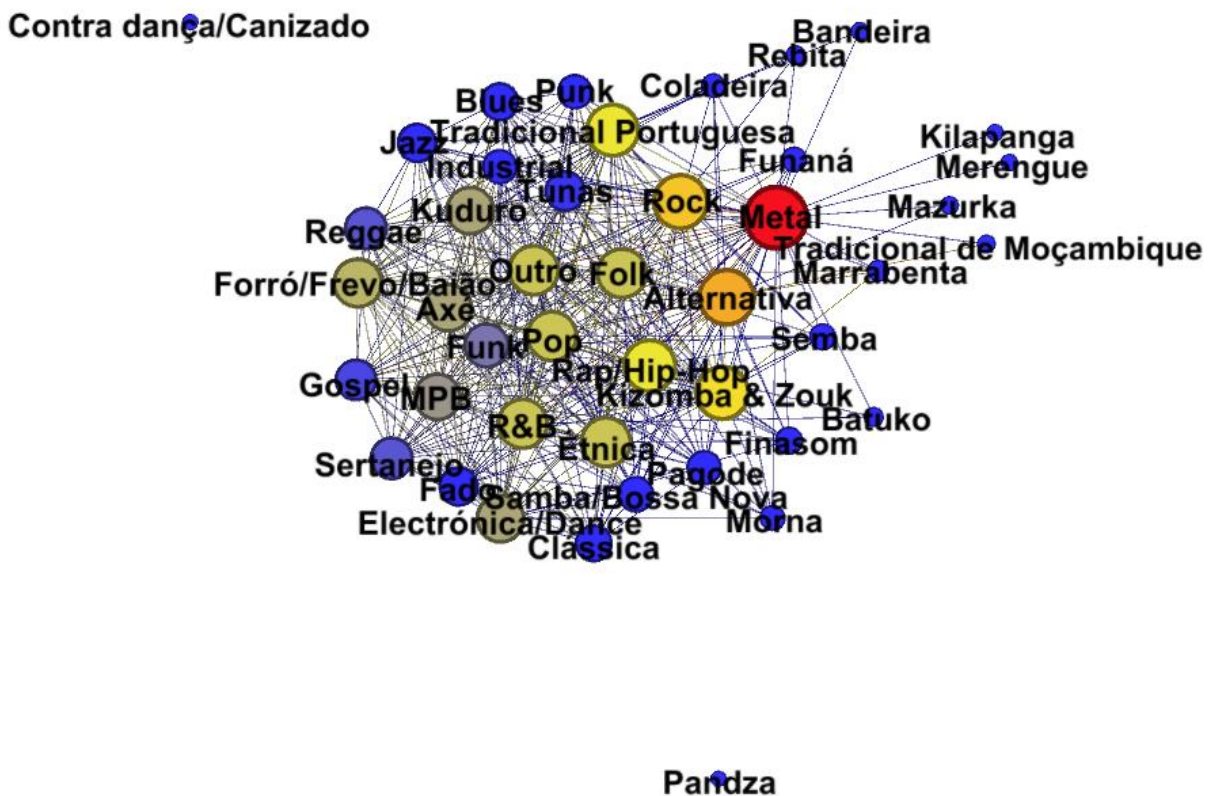


Figura 34 – Representação em rede de D_1

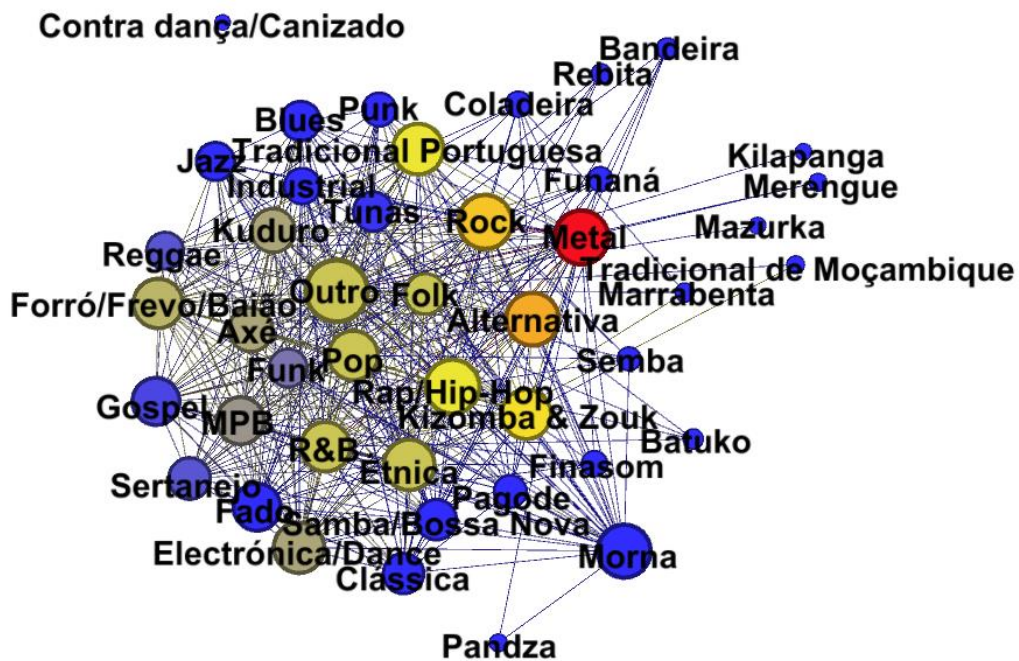


Figura 35 - Representação em rede de D_2

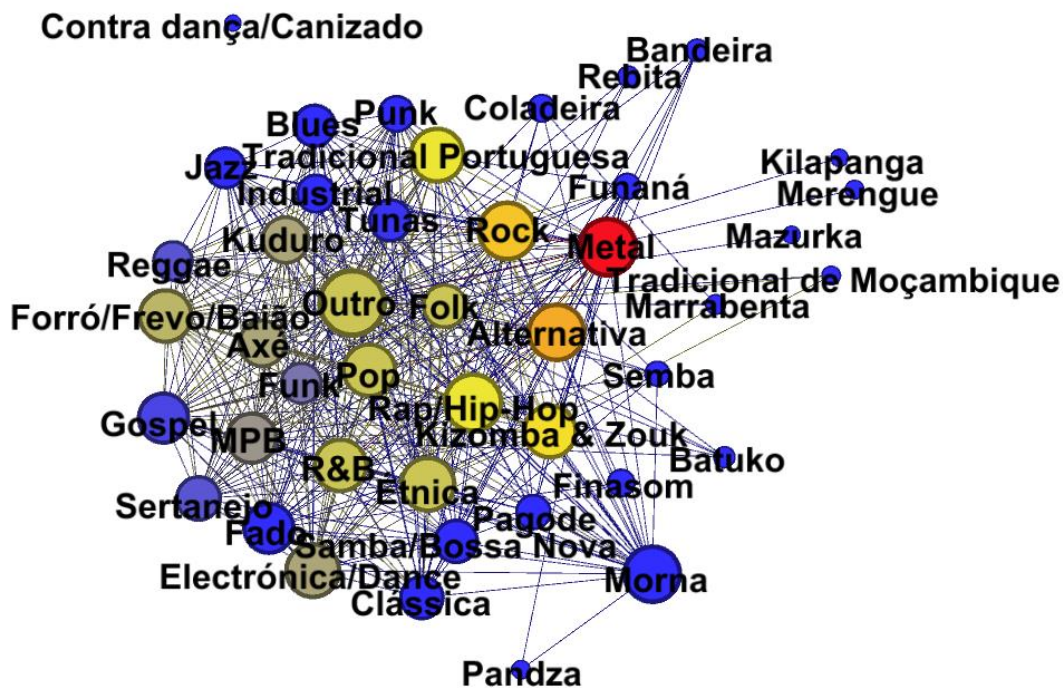


Figura 36 - Representação em rede de D_3

As medidas a nível de rede encontram-se na Tabela 12.

	D_1	D_2	D_3
Nº de utilizadores	7778	6317	5501
Nº Arestas	407	426	427
Média	44,56757	44,38967	40,00234
Dist. Média	12,30094	15,48216	13,41048
Densidade	0,430233	0,450317	0,451374
Transitividade gm	0,931978	0,930788	0,937166
Transitividade bi	0,762517	0,776214	0,802998

Tabela 12 - Tabela com as medidas a nível de rede para aparecimento/fortalecimento no método Janela Deslizante referentes a D_1 , D_2 e D_3

Na primeira linha da Tabela 12 podemos verificar que houve uma diminuição do número de novos utilizadores. Entre novembro de 2011 e janeiro de 2012 registaram-se apenas 5501 novos utilizadores, o que representa cerca de 70% do número de novos utilizadores registados entre setembro e novembro de 2011. Isto pode indicar um certo desinteresse do público por esta Rede Social.

Verifica-se que na terceira janela (D_3) existem mais 20 arestas do que na primeira, 19 das quais surgem na segunda janela. Isto poderá indicar que no período entre outubro e dezembro de 2011 houve uma abertura a novas preferências de estilos o que resultou em novas combinações entre vértices. Este facto pode estar ligado à entrada de novos utilizadores.

A média das ponderações das arestas (soma das ponderações pelo número de ligações) é mais baixa em D_3 . Neste período o número de novos utilizadores é menor que no período anterior mas o número de arestas apenas difere em uma unidade. Se considerarmos que as mudanças mais significativas na rede são provocadas por novos utilizadores, este argumento por si só explica esta mudança. Outra possível explicação é o facto de os novos utilizadores no período referente a esta janela terem os gostos musicais mais específicos e desta forma escolherem menos estilos musicais como favoritos. Esta situação traduz-se em menos contributos para a ponderação de arestas e consequente diminuição da média.

A distância média é a média dos valores da matriz de distâncias geodésicas entre os vértices. Relembremos que esta matriz é obtida a partir da matriz original após divisão

de todos os elementos pela média, seguida da inversão de cada um deles. Cada unidade refere-se à ponderação média de cada aresta. O valor de distância média aumenta do período entre setembro e novembro de 2011 (D_1), onde representa cerca de 28% da ponderação média das arestas, para o período entre outubro e dezembro de 2011 (D_2) onde representa aproximadamente 35%. O aumento do número de arestas é uma possível explicação para este facto. No segundo e terceiro períodos a percentagem em relação à média mantém-se em cerca de 34%.

A densidade da rede não é muito alta em nenhum dos períodos, sendo menor que 46%. Este resultado mostra que, nesta rede, mais de metade dos vértices não diretamente ligados.

Nas duas últimas linhas da Tabela 12 podemos comparar os valores de Transitividade (Coeficiente de *Clustering*) quando aplicada a distância geométrica (gm) na ponderação das arestas e quando é assumido que a matriz é binária (bi). Nas três janelas verificamos que usando a distância geométrica para os cálculos o valor desta medida é bastante mais elevado (aproximadamente 0.93 nas três janelas). Uma vez que é tão mais provável a criação de uma ligação que vai fechar um terceto aberto quanto maior a ponderação das duas arestas que unem os vértices desse terceto, esta medida é muito mais realista do que a calculada supondo a matriz binária. Note-se que, para o aumento desta medida na última linha da Tabela 12 é suficiente o aumento do número de arestas.

Em suma, estes resultados podem levar a crer que os utilizadores indicam menos estilos preferidos, no entanto, a variabilidade de preferência parece ser maior.

Medidas a nível de vértices:

A Tabela 13 mostra os vértices com o valor de excentricidade mais alta e mais baixa.

	D_1		D_2		D_3
Tradicional de Moçambique	66,85135	Tradicional de Moçambique	88,91377	Merengue	80,11788
Batuko	66,85135	Pandza	88,91377	Pandza	80,11788
Rebita	66,85135	Rebita	88,89524	Rebita	80,10346
Merengue	66,85135	Merengue	88,77934	Tradicional de Moçambique	60,46156
Mazurka	66,85135	Batuko	66,71893	Batuko	60,11671
Rap/Hip-Hop	44,65056	R&B	44,53612	R&B	40,17879
Rock	44,56757	Metal	44,52882	Rock	40,17042
Metal	44,56757	Outro	44,5241	Outro	40,11554
Tradicional Portuguesa	44,56757	Morna	44,5241	Morna	40,11554
		Rap/Hip-Hop	44,50557	Rap/Hip-Hop	40,10111

Tabela 13 - Excentricidade dos vértices para ganho/fortalecimento de ligações em D_1 , D_2 e D_3

À exceção de D_1 , onde não aparece o estilo “Pandza” (que na realidade não apresenta qualquer ligação nesta janela) os cinco vértices que apresentam o valor de excentricidade mais elevado não variam, sendo que em D_1 apresentam o mesmo valor. No segundo período, D_2 , verificamos um aumento considerável deste valor. Este facto pode ser em parte explicado pelas conexões criadas com o vértice “Pandza” que na primeira janela não apresentava qualquer ligação. De notar que “Tradicional de Moçambique” em D_3 deixa de pertencer ao grupo dos vértices com maior valor de excentricidade sendo a diminuição deste valor bastante considerável em relação ao diâmetro desta rede, isto pode indicar um gosto crescente dos utilizadores por este estilo de música.

Nas últimas linhas da Tabela 13 podemos observar os vértices com menor valor de excentricidade que se traduz numa menor facilidade em atingir os restantes vértices da rede. De salientar o facto desta medida ser mais baixa em D_3 o que indica encurtamento dos caminhos mais curtos ao longo do tempo quando relativizados à média de cada período. Em cada janela estes valores variam menos de 0,1 entre si o que indica que estes vértices atingem os restantes vértices da rede com a mesma facilidade.

Grau:

Em relação a esta medida vamos fazer a comparação de resultados quando $\alpha = 0$ em que o Grau corresponde ao número de vizinhos (Tabela 14) e quando $\alpha = 1.5$ em que é dada importância superior ao valor da ponderação das arestas incidentes no vértice do que à sua quantidade (Tabela 15).

alpha = 0	D_1		D_2		D_3
Metal	41	Outro	41	Outro	40
Alternativa	35	Rock	34	Rock	33
Rock	34	Rap/Hip-Hop	34	Morna	33
Kizomba & Zouk	33	Metal	34	Rap/Hip-Hop	32
Rap/Hip-Hop	32	Alternativa	33	Étnica	32
Mazurka	2	Tradicional de Moçambique	2	Rebita	3
Bandeira	2	Merengue	2	Pandza	2
Merengue	1	Pandza	2	Merengue	1
Kilapanga	1	Mazurka	1	Mazurka	1
		Kilapanga	1	Kilapanga	1

Tabela 14 – Vértices com maior e menor Grau na rede quando esta é considerada binária (alpha =0)

Na análise da Tabela 14 verificamos que em D_1 o estilo “Metal” aparece associado a mais estilos (41) do que nas Janelas seguintes. Nestas janelas o vértice mais associado a vértices representativos de outros estilos é “Outro” seguido de “Rock” com uma diferença considerável (7 vizinhos). Este resultado pode indicar uma perda de popularidade do vértice “Metal” nos períodos referentes a D_2 e D_3 entre os fãs de outros estilos ou pode indicar que os fãs de “Metal” nestes períodos apresentam gostos menos abrangentes. Em relação aos estilos com valor de Grau menor não se destacam diferenças significativas. Refere-se apenas que o estilo “Pandza” que não tinha vizinhos na primeira janela surge em duas combinações de duplas de estilos nas janelas seguintes.

alpha = 1.5	D_1		D_2		D_3
Rap/Hip-Hop	98711,02	Rap/Hip-Hop	113201	Rap/Hip-Hop	91019,26
R&B	38372,8	Outro	50161,13	Outro	59957,37
Alternativa	34133,61	Rock	33850,11	Rock	27693,45
Rock	27018,75	R&B	28276,93	Alternativa	20304,2
Outro	21856,5	Alternativa	26110,15	MPB	10928,35
Bandeira	3,674235	Mazurka	2,828427	Rebita	3
Rebita	3	Kilapanga	2,828427	Mazurka	2,828427
Merengue	2,828427	Tradicional de Moçambique	2	Kilapanga	2,828427
Kilapanga	2,828427	Merengue	2	Pandza	2
		Pandza	2	Merengue	1

Tabela 15 - Vértices com maior e menor Grau na rede quando é dada mais importância à ponderação da aresta do que ao número de vizinhos (alpha=1.5)

Na Tabela 15 verificamos que o estilo “Rap/Hip-Hop” se destaca com o valor de Grau mais elevado com grande diferença em relação ao segundo posicionado. Nas primeiras duas janelas, D_1 e D_2 , esta diferença é de mais do dobro. Isto mostra que, nos três períodos, ao contrário do que sugere a Tabela 14, este estilo musical apresenta um número mais alto de seguidores embora estes não se associem a uma variedade tão grande de estilos como o estilo “Metal” (Tabela 14) em D_1 e o vértice que representa outros estilos em D_2 e D_3 .

Proximidade (Closeness):

Os vértices com valores mais altos para esta medida são os que mais rapidamente atingem qualquer vértice na Rede e vice-versa. Nas Tabelas 16 e 17 mostram-se os resultados para os valores do parâmetro $\alpha=0$ e $\alpha=1.5$ à semelhança da análise do Grau.

alpha = 0	D_1		D_2		D_3
Metal	0,02439	Outro	0,023256	Outro	0,022727
Alternativa	0,021277	Rock	0,02	Rock	0,019608
Rock	0,020833	Rap/Hip-Hop	0,02	Morna	0,019608
Kizomba & Zouk	0,020408	Metal	0,02	Rap/Hip-Hop	0,019231
Rap/Hip-Hop	0,02	Alternativa	0,019608	Étnica	0,019231
Mazurka	0,0125	Merengue	0,012195	Merengue	0,011765
Bandeira	0,0125	Tradicional de Moçambique	0,012048	Mazurka	0,011765
Merengue	0,012346	Mazurka	0,011905	Kilapanga	0,011765
Kilapanga	0,012346	Kilapanga	0,011905	Finasom	0,011494
		Pandza	0,011236	Pandza	0,011111

Tabela 16 - Closeness quando a Rede é considerada binária

Na Tabela 16 observamos que o vértice “Metal” deixa de ser o vértice com maior valor de proximidade. Este vértice mostra, ao longo do tempo uma tendência para deixar de ser um vértice central na rede. Isto pode ser um indicador da diminuição do número de fãs deste estilo entre os utilizadores já na rede ou novos utilizadores. Em relação aos vértices com valor de proximidade mais baixos refere-se que “Bandeira” deixa de estar nas últimas quatro posições. Os restantes vértices não apresentam alterações significativas.

alpha = 1.5	D_1		D_2		D_3
Rap/Hip-Hop	0,000369295	Rap/Hip-Hop	0,000205	Rap/Hip-Hop	0,000258535
R&B	0,000369242	Outro	0,000205	Outro	0,000258534
Metal	0,000369177	R&B	0,000205	Rock	0,000258418
Alternativa	0,000369122	Rock	0,000205	R&B	0,000258369
Rock	0,000369063	Alternativa	0,000205	Morna	0,000258308
Mazurka	7,86703E-05	Kilapanga	6E-05	Kilapanga	0,000071
Kilapanga	7,86703E-05	Tradicional de Moçambique	2,62E-05	Tradicional de Moçambique	7,05353E-05
Bandeira	7,86425E-05	Rebita	2,62E-05	Rebita	3,03279E-05
Rebita	3,22597E-05	Merengue	2,62E-05	Merengue	3,03275E-05
		Pandza	2,62E-05	Pandza	3,03247E-05

Tabela 17 - Closeness quando é dada maior importância à ponderação das arestas

Na Tabela 17 salientamos que os vértices com valores mais altos nesta medida apresentam valores muito mais próximos entre si do que na Tabela 16. Isto acontece porque a ponderação das arestas vai mudar o comprimento dos caminhos. Em Anexo pode ser consultada a tabela completa onde se verifica que o estilo “Metal” apresenta o mesmo valor que os cinco vértices com valor mais elevado em D_2 (Tabela 17) e é muito próximo do valor mais elevado para esta medida em D_3 . Isto significa que ao contrário do que sugere a Tabela 16 este vértice continua a ser dos mais rápidos a atingir qualquer vértice na rede nos períodos seguintes. De notar também a maior discrepância entre os valores mais altos e mais baixos da Tabela 17 em relação à Tabela 16. Isto mostra que na realidade os vértices com valores mais baixos de *Closeness* vão demorar mais tempo a atingir qualquer outro vértice na Rede do que sugerido pela Tabela 16.

Intermediação (*Betweenness*):

O valor de Intermediação de um vértice mostra o quanto ele é importante na comunicação da rede. Quanto mais alto este valor mais caminhos mais curtos passam por este vértice. Nas Tabelas 18 e 19 estão os valores desta medida para $\alpha=0$ e $\alpha=1.5$.

alpha = 0	D_1		D_2		D_3
Metal	190,2578	Outro	187,0052	Outro	199,5879
Alternativa	54,33202	Metal	51,38855	Morna	44,95869
Tradicional Portuguesa	34,04511	Étnica	34,82488	Étnica	37,4669
Rock	32,13519	Rap/Hip-Hop	33,54331	Metal	28,04916
Kizomba & Zouk	23,36722	Morna	33,08855	Rap/Hip-Hop	27,1649
Merengue	0	Merengue	0	Merengue	0
Pandza	0	Pandza	0	Pandza	0
Mazurka	0	Mazurka	0	Mazurka	0
Bandeira	0	Bandeira	0	Bandeira	0
Kilapanga	0	Kilapanga	0	Kilapanga	0

Tabela 18 – Intermediação quando $\alpha=0$ (Rede binária)

Na análise da Tabela 18 verifica-se que existe uma grande diferença entre o vértice que apresenta o valor mais elevado para o segundo mais alto. De salientar que o estilo “Metal” deixa de ser o vértice com valor mais elevado no primeiro período para o segundo período o que faz sentido uma vez que, como já analisado na Tabela 14, o número de vizinhos baixa entre os dois períodos. Mais uma vez os resultados sugerem que o estilo “Metal” está a perder seguidores seja como preferência principal como preferência secundária. Em relação aos valores mais baixos nesta tabela refere-se apenas que são vértices que se encontram em posições mais periféricas da rede. Nesta tabela só se mostram cinco deles em cada janela mas em todas elas existem mais vértices com este valor igual a 0. Estes valores sugerem que ao longo do tempo estes vértices tendem a manter as mesmas posições uma vez que as apresentam nos 3 períodos de tempo diferentes.

alpha = 1.5	D_1		D_2		D_3
Rap/Hip-Hop	644	Rap/Hip-Hop	627	Rap/Hip-Hop	629
Metal	368	Outro	580	Outro	577
Kizomba & Zouk	239	Rock	120	Rock	155
Tradicional Portuguesa	158	MPB	81	MPB	120
Rock	119	Kizomba & Zouk	41	Alternativa	41
Merengue	0	Merengue	0	Merengue	0
Pandza	0	Pandza	0	Pandza	0
Mazurka	0	Mazurka	0	Mazurka	0
Bandeira	0	Bandeira	0	Bandeira	0
Kilapanga	0	Kilapanga	0	Kilapanga	0

Tabela 19 – Intermediação quando $\alpha=1.5$ (Rede ponderada)

Quando analisamos a Tabela 19 verificamos que o vértice mais usado em caminhos mais curtos representa o estilo “Rap/Hip-Hop”. Esta situação verifica-se nos três períodos de tempo o que mostra que este vértice é mais importante como intermediário entre outros do que se supunha pela análise da Tabela 18 nos três períodos de tempo. A diferença acentuada entre o valor desta medida nos dois vértices que apresentam os valores mais elevados em D_1 deixa de acontecer em D_2 e D_3 com a grande aproximação do vértice “Outro”. Este vértice tem uma interpretação diferente uma vez que pode representar estilos de música reconhecidos na rede ou não. O facto do vértice “Rap/Hip-Hop” apresentar um valor para esta medida superior reforça a sua importância nesta posição. A análise dos valores mais baixos nesta tabela é análoga à da Tabela 18.

Coeficiente de *Clustering* Local:

O Coeficiente de *Clustering* Local mostra o quanto um sub-grafo constituído pelos vizinhos de um vértice está próximo de ser completo, isto é, apresentar todas as arestas possíveis para um grafo com o mesmo número de vértices. No caso dos grafos ponderados este valor é calculado pela razão entre a soma da média geométrica dos tercetos fechados e a soma da média geométrica de todos os tercetos no sub-grafo correspondente a um certo vértice.

Este valor não é calculado para vértices que só apresentam um vizinho uma vez que o número de tercetos é nulo.

Qualquer vértice que apresente o valor 1 nesta medida tem a sua rede de vizinhos completa. Se um vértice está apenas ligado a dois vértices que se ligam entre si, então o seu coeficiente de *clustering* apresenta valor 1.

Na Tabela 20 está descrito quais os vértices que só apresentam um vizinho (em cima) e os que apresentam valor 1 (em baixo) nesta medida em cada janela.

D_1	D_2	D_3
Merengue Kilapanga	Mazurka Kilapanga	Merengue Mazurka Kilapanga

D_1	D_2	D_3
Tradicional de Moçambique Marrabenta Fado Funaná Samba/Bossa Nova Morna Semba Batuko Coladeira Finasom Rebita Mazurka Bandeira	Tradicional de Moçambique Marrabenta Semba Batuko Finasom Rebita Merengue Pandza Bandeira	Tradicional de Moçambique Marrabenta Semba Batuko Finasom Rebita Pandza Bandeira

Tabela 20 – Vértices para os quais não é calculada a medida do Coeficiente de *Clustering* (em cima) e vértices para os quais esta medida é um (em baixo) para as janelas D_1, D_2 e D_3

Com o aparecimento de novas arestas é de esperar que certos vértices deixem de ter todos os tercetos completos logo esta medida deixa de tomar o valor 1.

Nas Tabelas 21 e 22 encontram-se os resultados do Coeficiente de *Clustering* dos vértices que não estão na Tabela 20.

b_i	D_1		D_2		D_3
Industrial	0,991667	Jazz	0,994737	Tunas	0,994152
Punk	0,983333	Industrial	0,986928	Industrial	0,990476
Pagode	0,975	Samba/Bossa Nova	0,974026	Axé	0,985714
Clássica	0,96732	Sertanejo	0,968379	Jazz	0,965368
Jazz	0,957895	Tunas	0,957143	Clássica	0,964427
Tradicional Portuguesa	0,675403	Morna	0,6875	Rap/Hip-Hop	0,703629
Kizomba & Zouk	0,651515	Rock	0,663102	Rock	0,700758
Rock	0,631016	Rap/Hip-Hop	0,648841	Étnica	0,695565
Alternativa	0,593277	Metal	0,645276	Morna	0,683712
Metal	0,446341	Outro	0,467073	Outro	0,488462

Tabela 21 - Valores de Coeficiente de *Clustering* supondo a matriz binária

Os valores mais altos referem-se a vértices que apresentam vizinhanças com densidade mais elevada. Há maior probabilidade de criação de arestas nestes sub-grafos.

Nos vértices que apresentam valores mais baixos de Transitividade Local salientamos “Metal” em D_1 e “Outro” em D_2 e D_3 . Estes vértices apresentam um maior número de vizinhos (41, 41 e 40 respectivamente). Esta medida é baixa uma vez que existes muitos

estilos de música que não estão ligados entre si como sabemos pelo valor de densidade da rede que é baixa no três períodos de tempo.

gm	D_1		D_2		D_3
Industrial	0,996981	Jazz	0,998817	Tunas	0,996672
Punk	0,99663	Industrial	0,995818	Axé	0,994563
Gospel	0,992894	Sertanejo	0,995666	Industrial	0,993284
Sertanejo	0,992135	Punk	0,994217	Jazz	0,993245
Pagode	0,990477	Samba/Bossa Nova	0,993945	Clássica	0,991752
Alternativa	0,920051	Metal	0,913755	Rock	0,929024
Kizomba & Zouk	0,917804	Rock	0,906977	Morna	0,914481
Rock	0,911847	Étnica	0,903587	Étnica	0,896841
Tradicional Portuguesa	0,898136	Coladeira	0,858946	Outro	0,855043
Metal	0,821363	Outro	0,838419	Coladeira	0,845299

Tabela 22 - Valores de Coeficiente de *Clustering* usando a Distância Geométrica

Em relação aos vértices que apresentam um valor mais alto para o Coeficiente de *Clustering* a interpretação dada é a mesma que na Tabela 21.

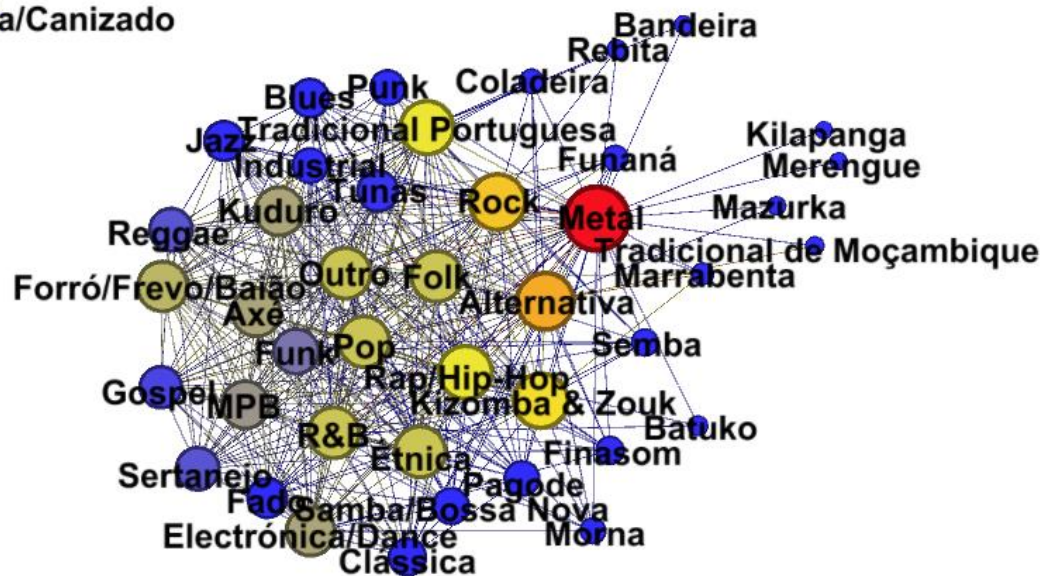
Destacamos o facto desta medida ser bastante mais alta quando usada a distância geométrica em relação à matriz binária. As ponderações levam a uma maior probabilidade de fecho de tercetos. É intuitivo pensar que para dois vértices que não se ligam entre si mas se ligam ambos a um terceiro a ligação entre eles é tão mais provável quanto mais alto for o valor da ponderação das arestas existentes neste terceto.

De um modo geral podemos dizer que a redução de cerca de 30% da quantidade de novos utilizadores da rede pode indiciar um desinteresse do público em geral. Em conjunto as medidas sugerem que as preferências dos utilizadores estão a criar novas combinações de estilos de música uma vez que no segundo período há criação de novas arestas. Uma possível explicação é a rede ter começado a ser utilizada por pessoas de uma região diferente de onde são originários certos tipos de música o que leva a uma maior diversificação das combinações de estilos. Faz-se ainda referência ao facto de quando os utilizadores escolhem nas suas preferências “Outros” se referirem a estilos que fazem parte da lista de estilos disponíveis mas simplesmente não os querem descrever a todos. Este é portanto um vértice que é de esperar que tenha uma posição central.

4.6.2. Janela Acumulada para aparecimento/aumento de ligações:

À semelhança do que acontece na janela deslizante também neste caso existem vértices isolados. Estes vértices representam os estilos “Contra dança/Canizado” e “Pandza” na primeira janela (A_1) e “Contra dança/Canizado” nas segunda (A_2) e terceira (A_3) janelas. Nas Figura 37 e 38 podemos ver a representação da rede em A_1 e A_2

Contra dança/Canizado



Pandza

Figura 37 – Janela Acumulada A_1

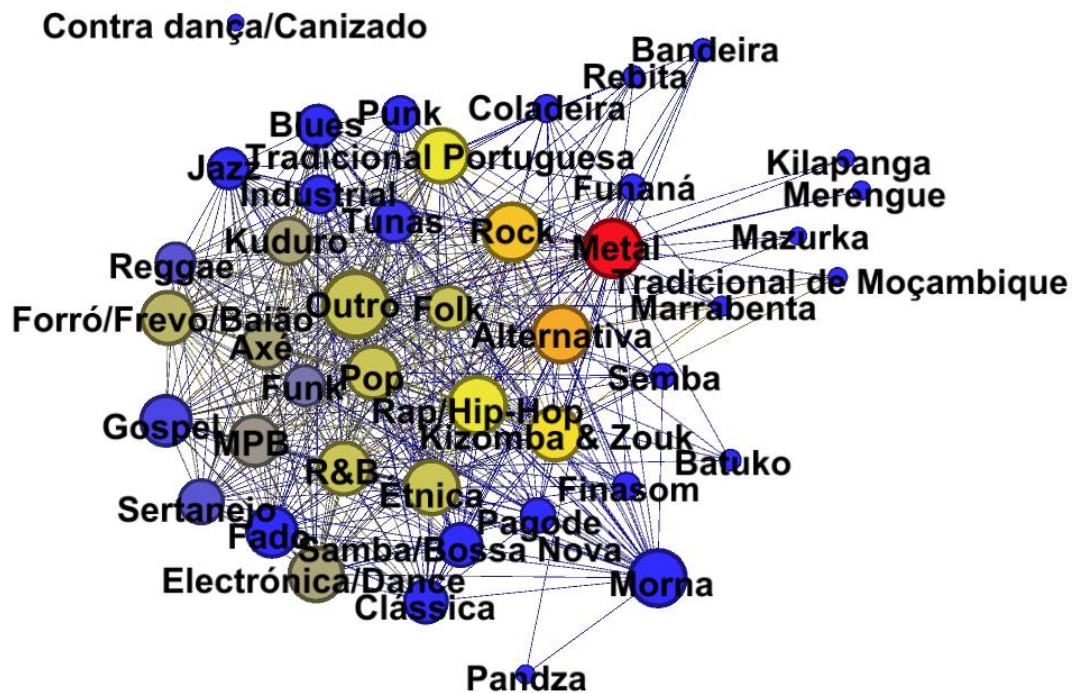


Figura 38 – Janela Acumulada A_2

A representação da terceira janela é semelhante à segunda pelo que não se mostra.

Na Tabela 23 resumem-se as medidas a nível de rede para estas três janelas.

	A_1	A_2	A_3
Nº de utilizadores	7778	11054	13279
Nº Arestas	407	481	498
Média	44,56757	62,89397	70,72289
Dist. Média	12,30094	17,78022	19,75373
Densidade	0,430233	0,508457	0,526427
Transitividade gm	0,931978	0,949266	0,958152
Transitividade bi	0,762517	0,795826	0,810944

Tabela 23 – Medidas a nível de rede para A_1 , A_2 e A_3

Com o aumento do número de utilizadores ao longo dos três períodos verifica-se o aumento do número de arestas. Este facto pode sugerir que está a haver maior diversidade de estilos escolhidos uma vez que aparecem diferentes combinações de estilos de música ao longo do tempo. Apesar disso a média aumenta, isto evidencia que

os pesos das arestas também estão a aumentar consideravelmente. Em combinação com a análise da janela deslizante verificamos que efetivamente existem muitas arestas sem atividade entre os diferentes momentos de tempo. Basta verificar que a diferença no número entre A_1 e A_2 é de 74. Estas arestas terão surgido entre novembro e dezembro de 2011 e ainda 17 arestas que terão sido criadas entre dezembro de 2011 e janeiro de 2012.

A distância geodésica média aumenta a cada período de tempo, isto indica que os vértices estão em média mais afastados entre si. Este afastamento está ligado a novas arestas que apresentam ponderações baixas. A densidade aumenta com a consideração de cada novo período, isto acontece uma vez que o número de arestas também aumenta. Os valores de Coeficiente de *Clustering* estão a aumentar o que é de esperar pelo aumento do número de arestas uma vez que aumentam o número de tercetos fechados. Mais uma vez o facto de usando a distância geométrica (Transitividade gm) o valor do Coeficiente de *Clustering* ser consideravelmente mais alto do que não considerando as ponderações (Transitividade bi) mostra que as ponderações das arestas não devem ser ignoradas.

A nível de vértices:

Nesta parte da análise vamos tomar apenas os resultados em que é dada mais importância à ponderação das arestas do que ao número de vizinhos de um vértice ($\alpha = 1.5$) uma vez que consideramos que estes resultados são mais representativos da realidade da rede.

Para valores de excentricidade, os vértices “Tradicional de Moçambique” e “Pandza” vão apresentar valores mais elevados em A_2 e A_3 com uma diferença considerável para a posição seguinte. Esta situação é de esperar especialmente no estilo “Pandza” uma vez que em A_1 não apresentava quaisquer ligações. Estes são os vértices mais difíceis de alcançar na rede. Nas posições mais baixas aparecem os vértices “Rap/Hip-Hop” e “Metal” embora estes valores estejam muito próximos dos apresentados por outros vértices. Estes vértices não variam muito nos três períodos o que significa que ao longo do tempo no intervalo de tempo entre setembro de 2011 e janeiro de 2012 os vértices

que mais rapidamente atingem qualquer outro vértice na rede não sofrem qualquer alteração. Em Anexo pode ser consultada uma tabela com estes valores.

Considerando valores obtidos para o Coeficiente de *Clustering* Local destacamos o vértice “Metal” com o valor mais baixo nas três Janelas. Uma vez que este vértice apresenta 41 vizinhos em A_1 e 42 vizinhos em A_2 e A_3 , a sua rede, além de apresentar muitos tercetos abertos, apresenta também valores das ponderações das arestas que formam tais tercetos muito discrepantes entre si.

A nível de Grau e Proximidade não se registam alterações significativas nas posições dos vértices representativos dos tipos de música nos três períodos de tempo. Em anexo podem ser consultadas tabelas com estes valores.

Já para valores de Intermediação (Tabela 24) destaca-se a grande diferença de valores nas três janelas entre os dois primeiros valores mais altos. Embora a análise da Janela deslizante sugira que o vértice “Rap/Hip-Hop” mostre uma tendência para deixar de ter o papel mais importante na ligação entre os restantes vértices uma vez que o vértice “Outro” apresenta um valor cada vez mais próxima nesta medida em D_2 e D_3 , esta análise sugere que as ligações mais antigas ainda o tornam o vértice mais importante na transmissão de informação na Rede.

alpha = 1.5	A_1		A_2		A_3
Rap/Hip-Hop	644	Rap/Hip-Hop	710	Rap/Hip-Hop	713,5
Metal	368	Outro	265	Outro	162,5
Kizomba & Zouk	239	Metal	129,5	Rock	150,5
Tradicional Portuguesa	158	Rock	118	Kizomba & Zouk	146
Rock	119	Kizomba & Zouk	111	Metal	126,5
Tunas	40	MPB	81	Tradicional Portuguesa	89
Alternativa	6	Tradicional Portuguesa	46,5	MPB	81
Outro	0	Tunas	41	Alternativa	41
R&B	0	R&B	0	Tunas	0

Tabela 24 – Valores de Intermediação para a Janelas Acumulada quando $\alpha=1,5$

4.7. Detecção de Comunidades

Nesta secção vão ser aplicados métodos de deteção de comunidades descritos no Capítulo 3 com o objetivo de obter uma partição inicial para aplicação de otimização por *Blockmodeling*.

Homogeneity Blockmodeling (Ziberna, 2006) é um método de detecção de comunidades que tem como objetivo a divisão da matriz de adjacência em blocos homogêneos. Tem como parâmetros de entrada o número de *clusters* e um modelo inicial. Por sucessivas permutações das linhas/colunas da matriz de adjacência este método vai procurar agrupar os vértices com o mesmo tipo de equivalência minimizando um erro.

Às partições resultantes será aplicado um método de detecção de eventos.

4.7.1. Aplicação de otimização pelo algoritmo Blockmodeling

No intuito de encontrar uma partição inicial para aplicação de otimização por *Blockmodeling* aplicaram-se às matrizes $D_i, i = 1, 2$ e $A_i, i = 1, 2$ e 3 , respetivamente as janelas deslizante e acumulada o algoritmo CNM, *Walktrap* e algoritmo de Girvan e Newman. Das partições obtidas as que indicaram melhor resultado a nível de valor de modularidade correspondiam às partições pelo algoritmo CNM. Resolveu-se por isso usar estas partições como partições iniciais para aplicação de *Blockmodeling*.

Ao resultado obtido foi aplicado o algoritmo *Homogeneity Blockmodeling* descrito anteriormente.

Na Figura 39 podemos observar os *clusters* (dentro de cada círculo) formados pelo algoritmo CNM na matriz D_1 (que por definição é igual a A_1).

Foram criadas 7 classes, duas delas, como seria de esperar, com os vértices que representam os estilos “Contra Dança/ Canzado” e “Pandza” que estão isolados. De resto, não se observa nestas classes nenhum factor em comum, como por exemplo os grupos juntarem músicas originárias do mesmo país.

Com a aplicação da otimização pelo método do *Blockmodeling* à partição encontrada anteriormente as diferenças são notórias e esperadas uma vez que vão ser procurados blocos com valores homogêneos. Na Figura 40 podemos observar o Grafo desta nova partição com os *clusters* devidamente numerados.

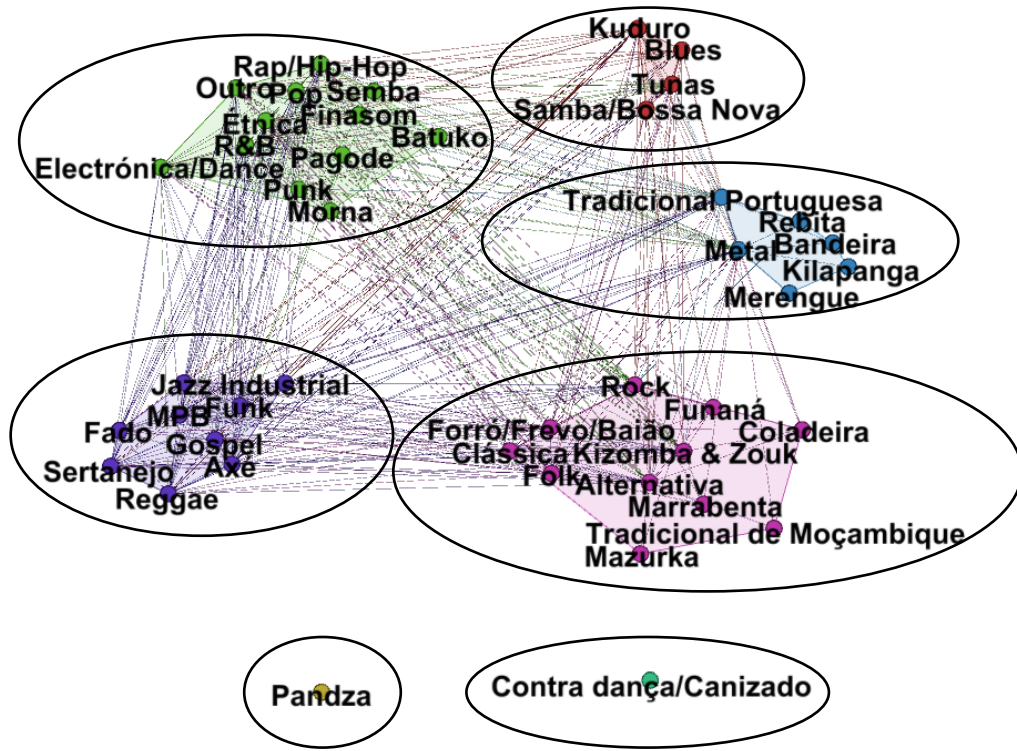


Figura 39 – Clusters do primeiro instante da Janela Deslizante D_1 formados pelo método *Fastgreedy*

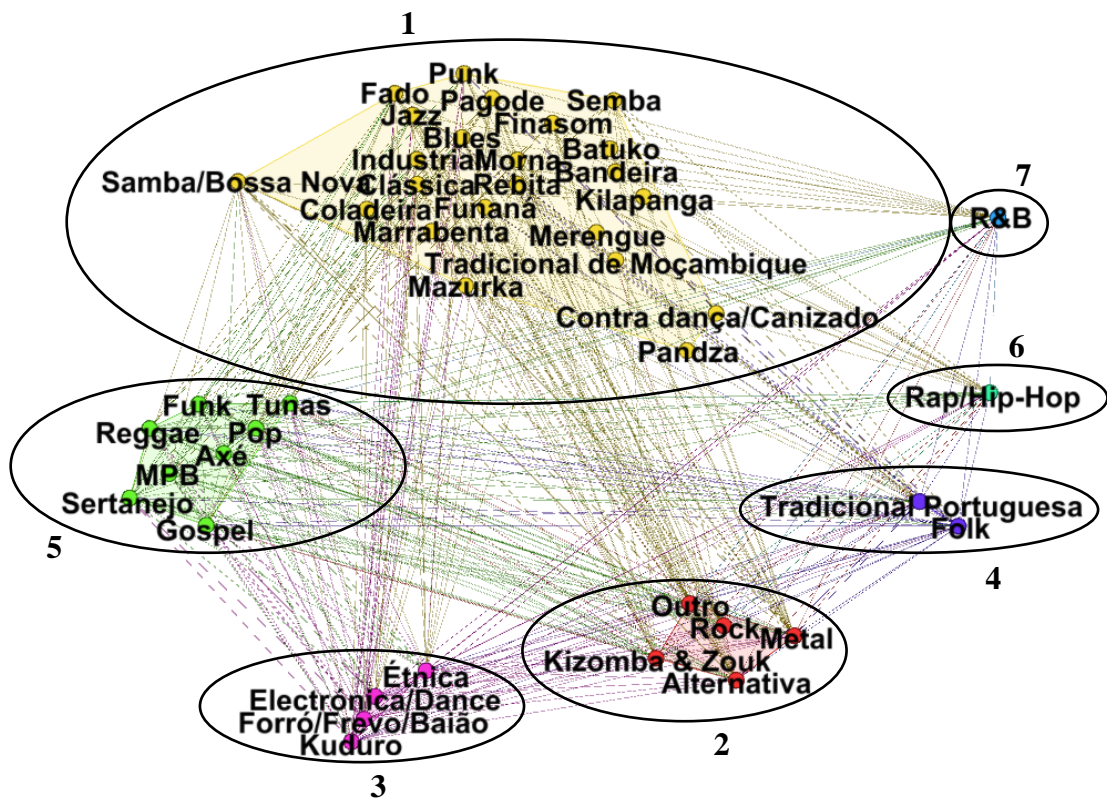


Figura 40 - Clusters formados pela aplicação de otimização por Blockmodeling a D_1

Começamos por verificar que os vértices “Rap/Hip-Hop” e “R&B” ficam isolados cada um constituindo um *Cluster*. Estes vértices apresentam valores na ponderação das suas arestas muito diferentes dos restantes na rede. Outra mudança notória é o facto de os vértices anteriormente isolados estarem agora integrados no mesmo grupo, onde se encontram os vértices com valores de ponderação de arestas mais baixos. Na matriz de adjacência da Figura 41 podem-se observar os valores das ligações dos blocos resultantes.

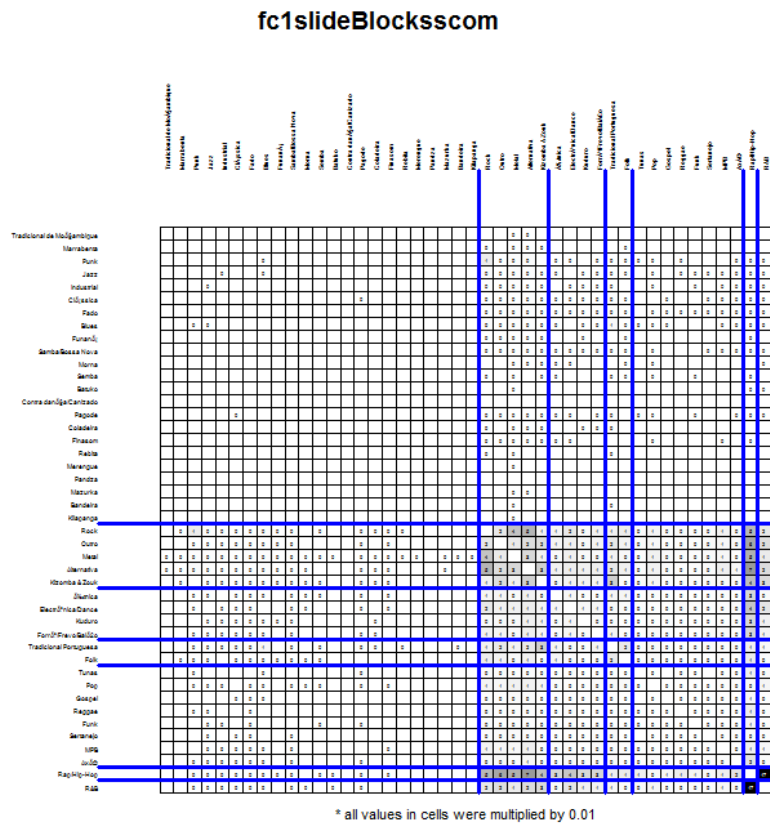


Figura 41 - Blocos formados na matriz D_1

4.7.2. Detecção de Eventos

De seguida procedeu-se à deteção de eventos. O método utilizado é baseado nos métodos descritos no Capítulo 3. A Tabela 25 mostra a maneira como os vértices de cada *cluster* da janela D_1 se distribuem pelos *clusters* de D_2 .

		clusters de D_2				
		cluster 1	cluster 2	cluster 3	cluster 4	cluster 5
clusters de D_1	cluster 1	0,869565	0,043478	0	0,086957	0
	cluster 2	0	0,4	0,4	0	0,2
	cluster 3	0	0,5	0	0,5	0
	cluster 4	0,5	0,5	0	0	0
	cluster 5	0,5	0,25	0	0,25	0
	cluster 6	0	0	0	0	1
	cluster 7	0	0	1	0	0

Tabela 25 – Distribuição dos *clusters* de D_1 pelos *clusters* de D_2

Por análise da Tabela 25 verificamos que os *cluster* candidatos a sobreviventes são os *clusters* 1, 6 e 7 de D_1 uma vez que apresentam uma percentagem de interseção com um *cluster* de D_2 igual ou superior a 60%. Os *clusters* 2, 3 e 4 foram divididos já que uma percentagem igual ou superior a 40% aparece em dois *clusters* distintos de D_2 . Não há absorção de *clusters* uma vez que não existem dois *clusters* em D_1 a contribuir com um mínimo de 60% cada um para um *cluster* em D_2 . O *cluster* 5 desaparece já que não se enquadra em nenhuma das situações anteriores.

Pela Tabela 26 podemos analisar as contribuições de cada *cluster* de D_1 para a formação de cada *cluster* de D_2 .

		clusters de D_2				
		cluster 1	cluster 2	cluster 3	cluster 4	cluster 5
clusters de D_1	cluster 1	0,8	0,125	0	0,333333	0
	cluster 2	0	0,25	0,666667	0	0,5
	cluster 3	0	0,25	0	0,333333	0
	cluster 4	0,04	0,125	0	0	0
	cluster 5	0,16	0,25	0	0,333333	0
	cluster 6	0	0	0	0	0,5
	cluster 7	0	0	0,333333	0	0

Tabela 26 – Contribuições de cada *cluster* de D_1 para cada *cluster* de D_2

Em relação ao *clusters* 1 de D_1 , podemos verificar que é realmente um sobrevivente uma vez que contribui para 80% do *cluster* 1 de D_2 . O *cluster* 6 de D_1 é também um sobrevivente visto que os seus vértices representam 50% do *cluster* 5 de D_2 pelo que vai continuar a denominar-se *cluster* 6. Já o *cluster* 7 de D_1 não sobrevive uma vez que representa apenas 33% do *cluster* 3 de D_2 . Os *clusters* 2, 3 e 4 nascem de D_1 para D_2 e passam a designar-se *clusters* 8, 9 e 10 respetivamente.

As Figuras 42 e 43 apresentam esquemas dos eventos que ocorreram entre D_1 e D_2 . Cada *cluster* de D_1 está representado por um círculo de fundo claro e cada *cluster* de D_2 está representado por um círculo de fundo escuro. O valor associado a cada círculo é o número do *cluster* e o valor associado a cada seta é o número de vértices que passaram do respetivo *cluster* de D_1 para o respetivo *cluster* em D_2 (para onde a seta aponta).

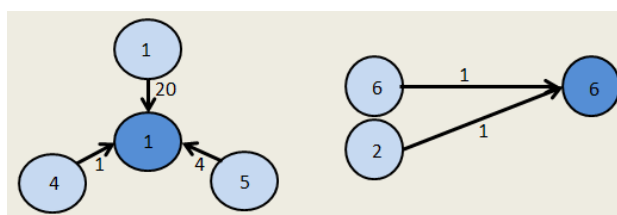


Figura 42 - *Clusters* sobreviventes entre D_1 e D_2

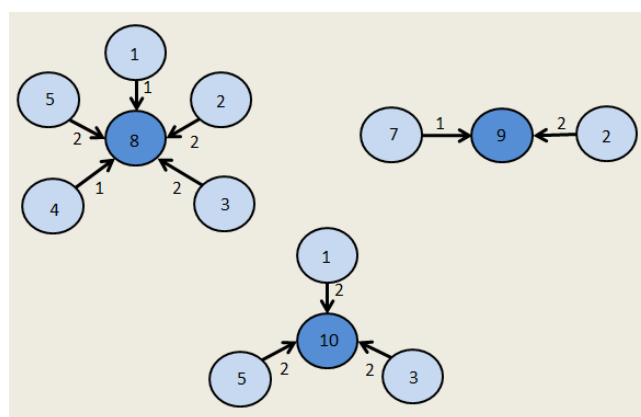


Figura 43 – Nascimento de *clusters* entre D_1 e D_2

O número de eventos ocorridos neste período de tempo leva-nos a crer que as ligações que se formaram são bastante diferentes do período correspondente a D_1 . O facto de

haver formação de 3 novos *clusters* mostra que os vértices estão organizados de maneira muito diferente. Em anexo podemos observar a rede formada em D_2 . Salientamos o vértice “Outro” que se reúne com o *cluster* constituído pelo vértice “Rap/Hip-Hop”. Este resultado vai ao encontro de resultados das análises anteriores. Os vértices “Alternativa”, “Rock” e “R&B” formam o *cluster* 9. Neste período as ligações formadas entre estes vértices e os restantes da rede estão mais próximas. O *cluster* 1, que sobrevive agrupa vértices que apresentam ligações baixas tanto entre si como com os restantes vértices na rede.

De uma forma análoga foram analisadas as mudanças entre D_2 e D_3 . As Figuras 44 e 45 esquematizam os resultados. Os círculos com fundo claro representam *clusters* de D_2 e os círculos com fundo escuro representam *clusters* de D_3 .

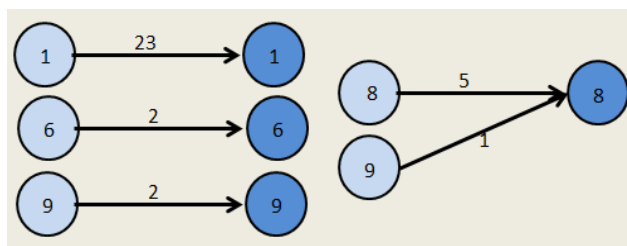


Figura 44 - *Clusters* sobreviventes entre D_2 e D_3

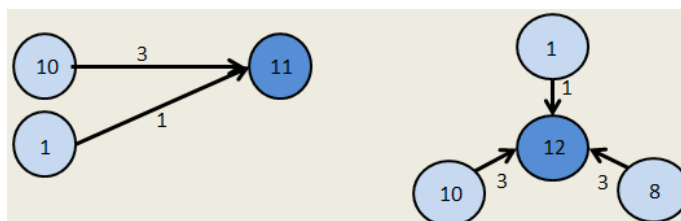


Figura 45 – Nascimento de *clusters* entre D_2 e D_3

Neste período as mudanças são bastante menos acentuadas. Quatro dos cinco *clusters* de D_2 sobrevivem em D_3 . Apenas o *cluster* 10 se divide. Este *cluster* agrupa os vértices “Forró/Frevo/Baião”, “Gospel”, “Punk”, “Étnica”, “Sertanejo” e “Fado”. Uma vez que este *cluster* se formou entre D_1 e D_2 , e entre D_2 e D_3 se dividiu leva a crer que estes vértices serão os que mais variam as ligações que criam com os restantes vértices ao longo do tempo.

De uma forma geral estes resultados mostram que existiu uma maior variação nas preferências dos utilizadores entre setembro e dezembro de 2011 do que entre outubro de 2011 e janeiro de 2012. Os vértices “Rap/Hip-Hop” e “Outro” aproximam-se no tipo de ligações assim como os vértices “Alternativa” e “Rock”. O cluster 1, cujos vértices apresentam valores mais baixos nas ponderações das arestas mostram tendência para se manter na mesma forma.

Em anexo podem ser consultadas as tabelas de contribuições de *clusters* de D_2 para D_3 assim como a representação completa da rede nestes períodos.

4.8. Aplicação do *Blockmodeling* à partição obtida por *Clustering Hierárquico*

Aplicou-se o algoritmo de *clustering* hierárquico à matriz de adjacência D_1 , usando a distância Euclidiana e o critério de agregação da média. O dendrograma obtido está na Figura 46:

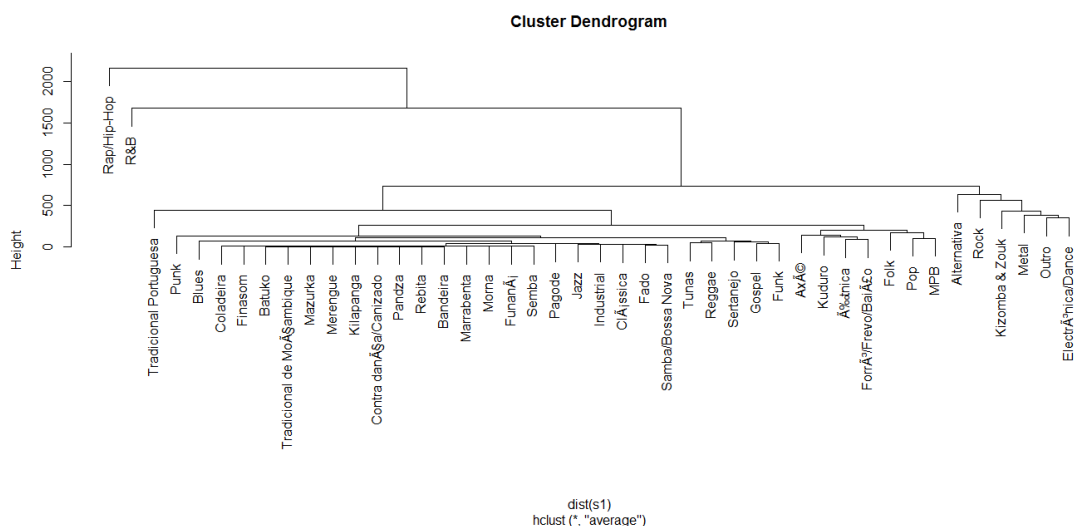


Figura 46 – Dendrograma obtido pela aplicação de *clustering* hierárquico a D_1

Foi feito um corte para 4 classes. Desta forma os vértices “Rap/Hip-Hop” e “R&B” formam um *cluster* cada um. A esta partição aplicou-se a otimização pelo método

Blockmodeling. O resultado para a primeira janela deslizante, D_1 (e acumulada, A_1) está representado na Figura 47.

Podemos verificar que localmente a partição inicial é a partição ótima encontrada pelo *Blockmodeling*. De seguida usa-se esta partição como partição inicial para otimização por *Blockmodeling* na matriz de adjacência da segunda Janela Deslizante, D_2 . O resultado está representado na Figura 48.

Na Figura 48 podemos observar que houve duas mudanças relacionadas com os vértices “Morna” e “Outro”. O vértice “Outro” foi agrupado com o estilo “R&B” o que mostra uma aproximação dos tipos de ligação dos dois estilos no período de outubro a dezembro de 2011. Os tipos de ligações do estilo “Morna” estão agora mais próximos dos estilos “Metal”, “Alternativa”, “Kizomba & Zouk”, “Rock” e “Eletrónica/Dance”. O estilo musical “Rap/Hip-Hop” continua isolado num grupo mostrando que neste período de tempo as ligações que este cria continuam a ser bastante diferentes do resto dos vértices da rede.

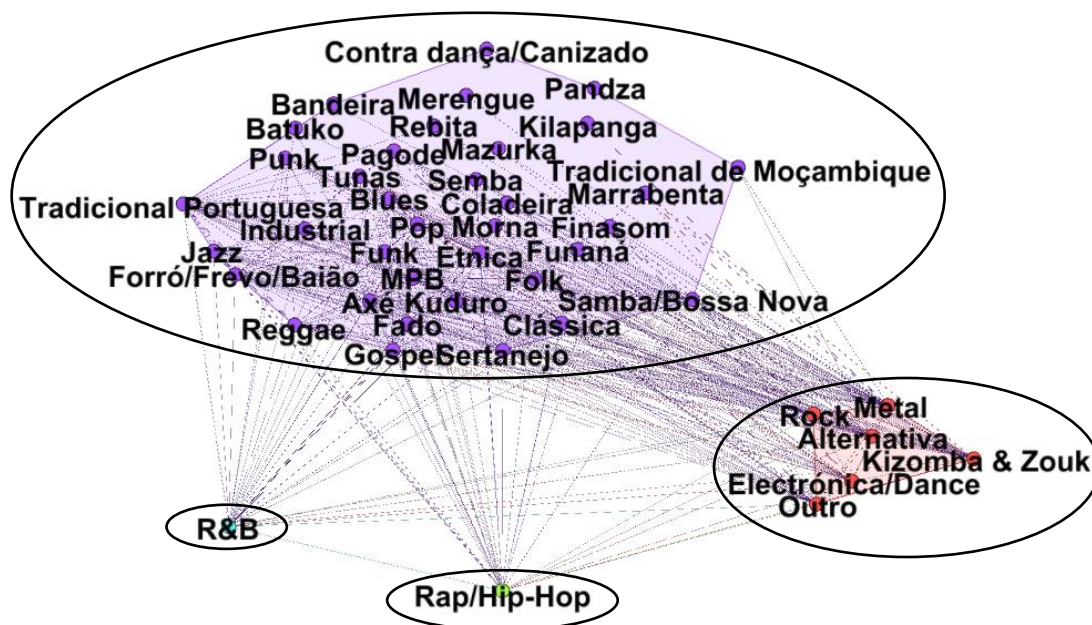


Figura 47 – Clusters formados pela aplicação de otimização por *Blockmodeling* a D_1

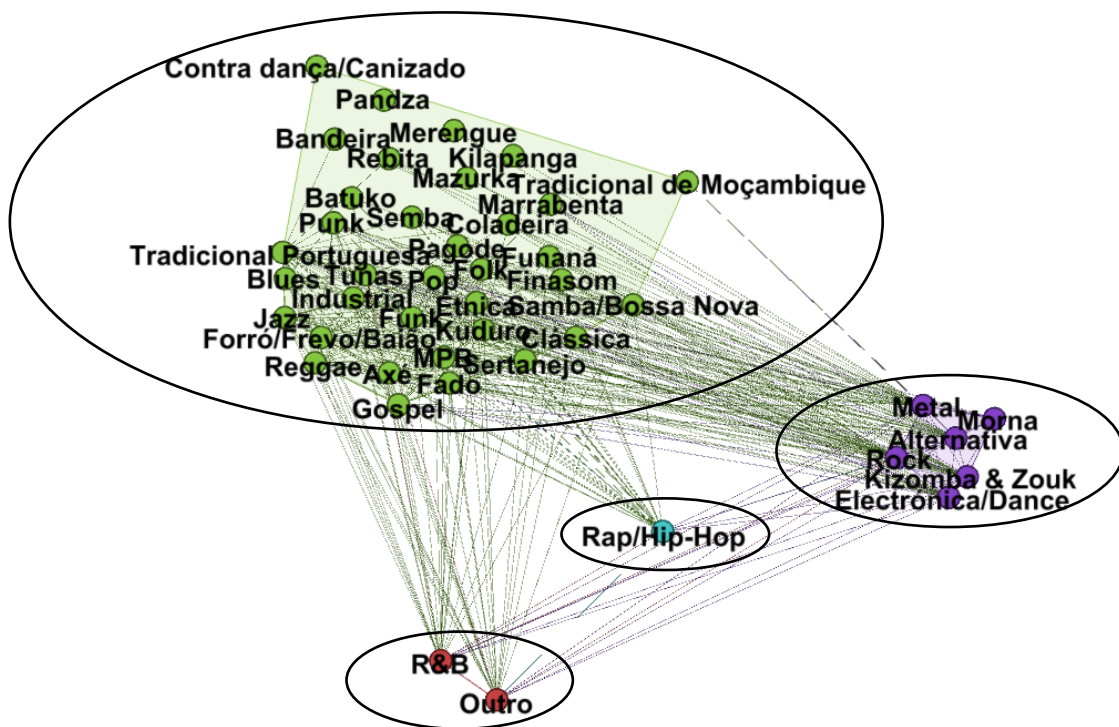


Figura 48 – Clusters formados pela aplicação de otimização por Blockmodeling a D_2

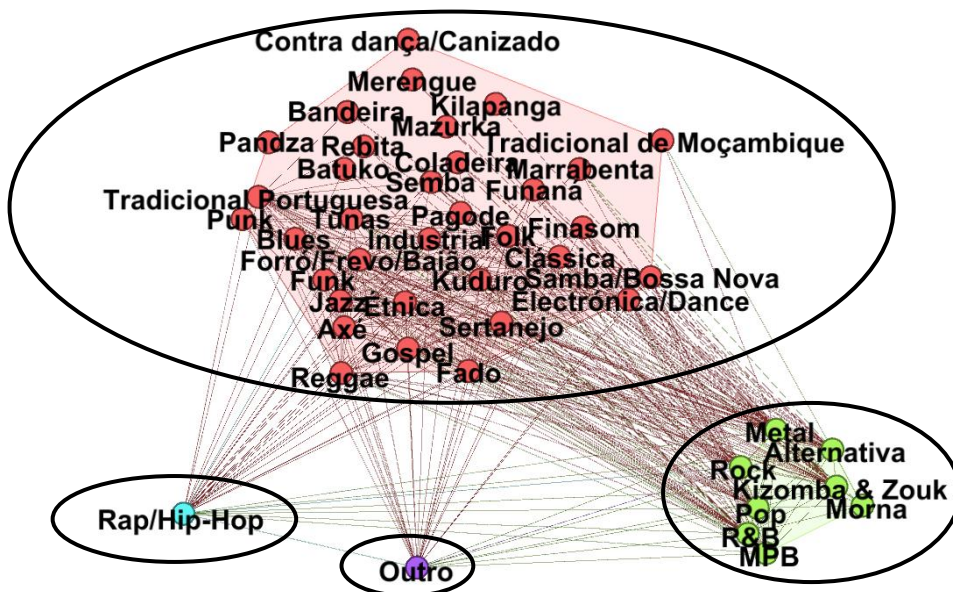


Figura 49 – Clusters formados pela aplicação de otimização por Blockmodeling a D_3

Para a terceira janela deslizante, D_3 (Figura 49) note-se que o vértice “Outro” está pela primeira vez sozinho num *cluster*. Numa análise mais cuidada verificou-se que a soma das ponderações das arestas deste vértice aumentaram em cerca de 46% em relação ao valor do período entre setembro e novembro de 2011. Isto reflete um grande aumento referente ao mês de novembro de 2011. O estilo “R&B” foi agora para o *cluster* de vértices como o “Rock” ou “Morna”. As ponderações associadas às arestas deste vértice têm vindo a diminuir desde a primeira janela o que indica que os utilizadores da rede poderão estar a perder interesse neste estilo. O estilo “Rap/Hip-Hop” continua a ser o mais referido nas preferências.

Aplicação do *Blockmodeling* à Janela Acumulada

A janela acumulada A_1 coincide por definição com a janela deslizante D_1 . A Figura 50 mostra o resultado da aplicação do algoritmo *Blockmodeling* a A_2 usando a partição resultante da aplicação do mesmo algoritmo a A_1 .

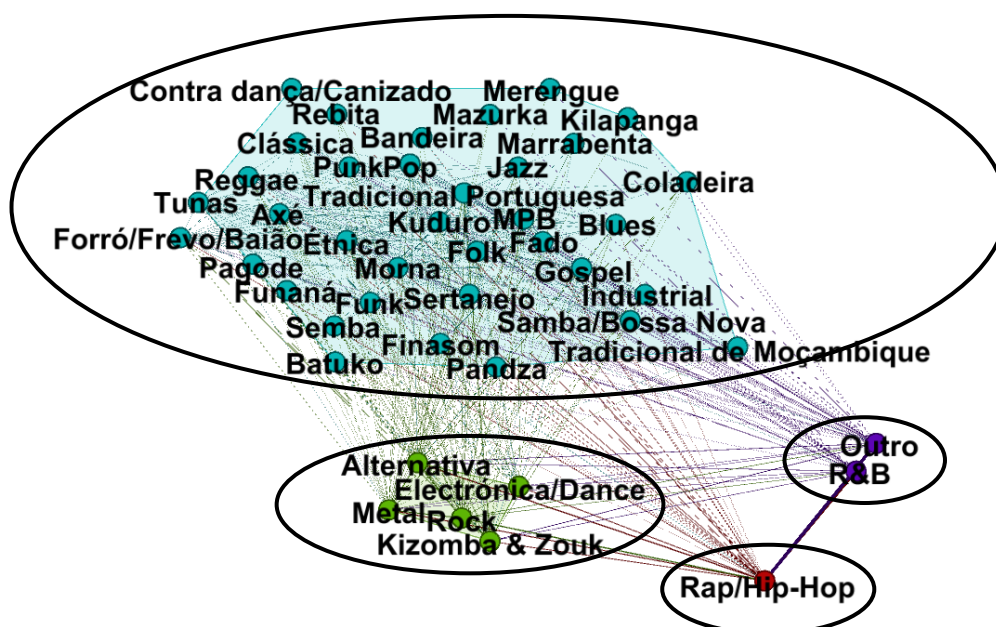


Figura 50 – Segunda Janela Acumulada

Podemos verificar que a única alteração é o vértice “Outro” ter mudado para o *cluster* do estilo “R&B” à semelhança do que aconteceu para D_2 . As ligações criadas no mês de dezembro aproximam estes dois vértices. O vértice “Rap/Hip-Hop” continua isolado. Este resultado mostra que este vértice continua a ser o que mais sucesso tem entre as preferências dos utilizadores.

Da segunda para a terceira janela acumulada não se verifica qualquer alteração.

Em suma, pela análise das janelas deslizantes e acumuladas quando aplicada a otimização por Blockmodeling à partição resultante da janela anterior podemos dizer que a rede não apresenta alterações significativas. O estilo de música “Rap/Hip-Hop” é a grande preferência dos utilizadores desta rede. O vértice “Outro” é o que mais se aproxima deste a nível de ligações idênticas com os restantes vértices. Existe no entanto uma ambiguidade na interpretação do vértice “Outro” uma vez que este se pode referir a estilos de música com ou sem representação nesta rede. Esta ambiguidade mais uma vez reforça o sucesso do estilo “Rap/ Hip-Hop”. Embora em D_3 possa haver a indicação de uma baixa nas preferências dos utilizadores para o estilo “R&B” tal não é conclusivo.

Capítulo 5

Conclusões

Este trabalho teve como objetivo analisar a rede de preferências musicais da rede social Palco Principal. Cada vértice desta rede representa um estilo musical e cada aresta é ponderada com o número de pessoas que apresenta simultaneamente nas suas preferências os dois estilos que a une.

Na análise das métricas a nível de vértice verificou-se que o estilo com mais sucesso no período de setembro de 2011 a janeiro de 2012 é “Rap/ Hip-Hop”. Estilos como “Rock”, “Alternativa”, “R&B”, “Eletrónica/ Dance”, “Kizomba e Zouk” ou “Metal” são também estilos muito escolhidos nas preferências dos utilizadores o que os tornam vértices mais centrais na rede. Os restantes vértices apresentam em geral ligações com ponderações baixas e um número de vizinhos reduzido.

Estes resultados também se refletem na parte de deteção de comunidades após aplicação de otimização por *Blockmodeling*. O vértice mais próximo de “Rap/ Hip-Hop” é o vértice “Outro”. Este vértice pode ser interpretado como preferência tanto por estilos representados por outros vértices ou estilos não referidos na rede. O facto de “Outro” ser o único vértice que se aproxima de “Rap/Hip-Hop” reforça o seu sucesso. Esta análise torna mais notório o facto de mais de metade dos vértices da rede apresentarem ponderações baixas para as suas ligações. Em geral estes vértices aparecem agrupados.

A nível de dinâmica podemos dizer que as ligações que surgem na rede nos diferentes intervalos de tempo não diferem. Tanto pela análise das métricas como pela análise da evolução de *clusters* podemos concluir que mais de metade dos vértices mantém um número baixo de vizinhos assim como baixas ponderações nas suas arestas. Existe um conjunto de vértices do qual fazem parte “Rock”, “Alternativa”, “R&B”, “Eletrónica/ Dance”, “Kizomba e Zouk” ou “Metal” que são bastante referidos nas preferências dos

utilizadores. É neste grupo que se notam algumas diferenças a nível de dinâmica que não são no entanto significativas. Por fim, mais uma vez aparece o vértice “Rap/Hip-Hop” cujas ligações apenas se aproximam do vértice “Outro”

Em relação às ferramentas utilizadas refere-se a grande utilidade da *package* *tnet* do R que se revelou de grande importância para o estudo de redes ponderadas e conexas definidas com matrizes de distâncias. Refere-se também a *package* *Blockmodeling*, a análise de rede com o uso deste algoritmo permite uma perspetiva diferente de deteção de comunidades já que agrupa vértices que apresentam o mesmo tipo de ligações entre si e com o resto da rede.

5.1. Trabalho futuro

Como trabalho futuro propõe-se que, na parte de deteção de eventos seja incluída a ponderação das arestas. Um grupo pode não ser definido pela quantidade de indivíduos mas pelas relações que estas apresentam entre si. Por outras palavras, se um grupo de 20 pessoas passa a ser de 5, não implica que este grupo tenha deixado de existir. Estas 5 pessoas podem ser o núcleo do grupo pelo que o grupo não desapareceu. Seria interessante verificar as diferenças entre o trabalho efetuado e esta perspetiva.

Bibliografia

- Aldous, J.M. and Wilson, R. J. (2000). *Graphs and Applications: an introductory approach*. Springer, Great Britain.
- Batagelj, V. (1997). Notes on Blockmodeling. *Social Networks* 19:143-155.
- Bondy, J. A. (1976). *Graph Theory with Application*. Elsevier Science Ltd.
- Choobdar, S.; Ribeiro, P; Silva, F, 2012, Event Detection in Evolving Networks, conferência CASoN, 26-32.
- Clauset, A.; Newman, M.E.J. and Moore, C. (2004), Finding community structure in very large networks, *Phys. Rev. E* 70, 066111.
- Doreian, P.; Batagelj, V. and Ferligoj, A. (2005). *Generalized Blockmodeling*. Cambridge University Press,
- Fortunato, S. (2010). "Community detection in graphs". *Physics Reports* 486, 75 - 174, *Physics Reports*.
- Gama, J., 2012, Data Stream Mining: the Bounded Rationality, *Informatica* 37 (2013) 21–25
- Granovetter, M. (1973). The strength of weak Ties. *American Journal of Sociology*, 78(6):1360-1380.
- Han, J., Kamber, M. and Pei, J. (2012), *Data Mining: Concepts and Techniques*, third edition.
- Harary, F., Norman, R. and Cartwright, D., 1965, *Structural Models*. New York: Wiley.
- Newman, M. (2003). The structure and function of Complex Networks. *Society for Industrial and Applied Mathematics* 45(2):167-256.
- Newman, M.E.J. (2003), Fast algorithm for detecting community structure in networks, *Phys. Rev. E* 69, 066133.
- Newman, M.E.J. (2004), Analysis of weighted networks, *Phys. Rev. E* 70, 056131.
- Newman, M. E. J. and Girvan, M., 2004, Finding and Evaluating community structure in Networks, *Phys. Rev. E* 69, 026113
- Nooy, W.; Mvar, A. and Batagelj, V. (2005). *Exploratory Social Network Analysis with Pajek*. Cambridge University Press, Cambridge.
- Oliveira, M., Gama, J. (2012). An Overview of Graphs Social Network Analysis. *WIRES Data Mining Knowledge Discovery* 2012, 2: 99–115 doi: 10.1002/widm.1048.

Opsahl, T.; Agneessens, F. and Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks* 32 (2010) 245–251, DOI: 10.1016/j.socnet.2010.03.006

Opsahl, T. and Panzarasa, P.; (2009). *Clustering* in weighted networks. *Social Networks* 31 (2009), Pages 155–163, DOI: 10.1016/j.socnet.2009.02.002.

Opsahl, T., “Shortest Paths in Weighted Networks” from <http://toreopsahl.com/tnet/weighted-networks/shortest-paths/> acessado pela última vez a 30 de Junho de 2014

Ponts, P., Latapy, M. (2005), Computing communities in large networks using random walks, *Journal of Graph Algorithms and Applications*, vol. 10, no. 2, pp. 191–218.

Sandifer, C. (2007). *The Early Mathematics of Leonhard Euler*. The Mathematical Association of America.

Spiliopoulou, M., Ntoutsis, I., Theodoridis, Y. and Schult, R., 2006, MONIC – Modeling and Monitoring Clusters Transitions, Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining Pages 706-711

Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge.

Watts, DJ., Strogatz, SH., 1998, Collective dynamics of small-world networks, *Nature* 1998, 393:440-442.

Ziberna, A, (2007), Generalized blockmodeling of valued networks, *Social Networks* 29 (2007) 105–126

Anexo 1 – Tabelas com valores de medidas a nível de vértice em A_1 , A_2 e A_3 considerando $\alpha = 1.5$

alpha = 1.5	A_1		A_2		A_3
Trad. de Moçambique	66,85135	Rebita	125,7879	Rebita	141,4458
Batuko	66,85135	Pandza	125,7879	Pandza	141,4458
Rebita	66,85135	Trad. de Moçambique	94,34096	Trad. de Moçambique	106,0843
Merengue	66,85135	Batuko	94,34096	Batuko	106,0843
Mazurka	66,85135	Merengue	94,34096	Merengue	106,0843
<hr/>					
Alternativa	44,65565	Alternativa	63,04965	Alternativa	70,88254
Rap/Hip-Hop	44,65056	R&B	63,01538	R&B	70,85007
Rock	44,56757	Outro	63,00935	Outro	70,83957
Metal	44,56757	Rock	63,00112	Rock	70,8322
Tradicional Portuguesa	44,56757	Rap/Hip-Hop	62,98646	Rap/Hip-Hop	70,81859
		Metal	62,89397	Metal	70,72289

Tabela 27 – Valores de Excentricidade em A_1 , A_2 e A_3 com $\alpha = 1.5$

alpha = 1.5	A_1		A_2		A_3
Rap/Hip-Hop	98711,02	Rap/Hip-Hop	209080,9	Rap/Hip-Hop	256491,9
R&B	38372,8	Outro	78280,23	Outro	104865,2
Alternativa	34133,61	Rock	62052,74	Rock	75689,17
Rock	27018,75	R&B	58041,5	Alternativa	72242,19
Outro	21856,5	Alternativa	57236,8	R&B	64511
<hr/>					
Bandeira	3,674235	Rebita	6	Mazurka	6,454972
Rebita	3	Kilapanga	5,656854	Rebita	6
Merengue	2,828427	Trad. de Moçambique	5,59017	Kilapanga	5,656854
Kilapanga	2,828427	Merengue	3,674235	Merengue	3,674235
		Pandza	2	Pandza	2

Tabela 28 - Valores de Grau em A_1 , A_2 e A_3 com $\alpha = 1.5$

alpha = 1.5	A_1		A_2		A_3
Rap/Hip-Hop	0,000369	Rap/Hip-Hop	0,000175	Rap/Hip-Hop	0,000147
R&B	0,000369	Outro	0,000175	Outro	0,000147
Metal	0,000369	R&B	0,000175	R&B	0,000147
Alternativa	0,000369	Rock	0,000175	Rock	0,000147
Rock	0,000369	Alternativa	0,000175	Alternativa	0,000147
Outro	0,000369	Metal	0,000175	Metal	0,000147
<hr/>					
Mazurka	7,87E-05	Kilapanga	4,13E-05	Kilapanga	3,45E-05
Kilapanga	7,87E-05	Trad. de Moçambique	4,13E-05	Trad. de Moçambique	3,45E-05
Bandeira	7,86E-05	Merengue	4,13E-05	Merengue	3,45E-05
Rebita	3,23E-05	Rebita	1,72E-05	Rebita	1,44E-05
		Pandza	1,72E-05	Pandza	1,44E-05

Tabela 29 - Valores de Proximidade em A_1 , A_2 e A_3 com $\alpha = 1.5$

Anexo 2 – Tabelas de contribuições de vértices entre as janelas D_2 e D_3 e representação em rede das mesmas janelas

		clusters de D_3					
		blofc31	blofc32	blofc33	blofc34	blofc35	blofc36
clusters de D_2	cluster 1	0,92	0	0,04	0,04	0	0
	cluster 8	0	0,625	0	0,375	0	0
	cluster 9	0	0,333333	0	0	0,666667	0
	cluster 10	0	0	0,5	0,5	0	0
	cluster 6	0	0	0	0	0	1

Tabela 30 – Distribuição dos clusters de D_2 pelos clusters de D_3

		clusters de D_3					
		blofc31	blofc32	blofc33	blofc34	blofc35	blofc36
clusters de D_2	cluster 1	1	0	0,25	0,142857	0	0
	cluster 8	0	0,833333	0	0,428571	0	0
	cluster 9	0	0,166667	0	0	1	0
	cluster 10	0	0	0,75	0,428571	0	0
	cluster 6	0	0	0	0	0	1

Tabela 31 – Contribuições de cada cluster de D_2 para cada cluster de D_3

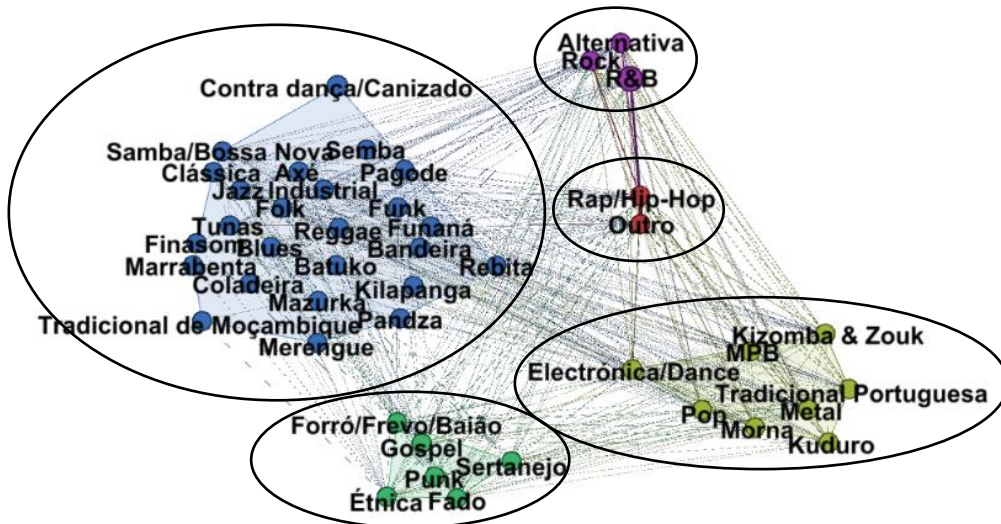


Figura 51 – Partição do Grafo da janela D_2 através da aplicação do *blockmodeling*

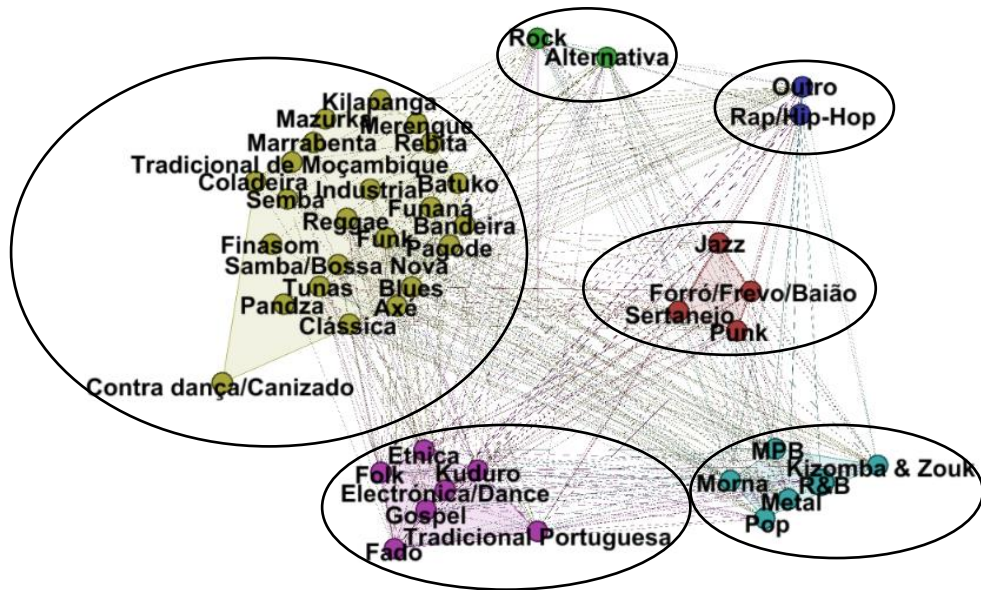


Figura 52 – Partição do Grafo da janela D_3 através da aplicação do *blockmodeling*