



**Grafos e redes sociais:
Cliques e relaxações de cliques;
Medidas de centralidade**

Laura Pinguicha Ferreira

Mestrado em Estatística e Investigação Operacional
Especialização em Investigação Operacional

Dissertação orientada por:
Professor Doutor Luís Eduardo Neves Gouveia

Resumo

O módulo de grafos da disciplina “Introdução aos Modelos Matemáticos” estuda a aplicação de grafos à análise de redes sociais. Nesta perspectiva, os nodos de um grafo representam pessoas e uma aresta entre dois nodos indica que as pessoas correspondentes se conhecem. A análise de grupos de indivíduos e das relações entre eles pode ser particularmente útil quando estes indivíduos são “indesejáveis”, como é o caso de criminosos ou terroristas. Poderá também ser útil na construção de estratégias de marketing que pretendem atingir pessoas com determinadas características. Esse processo pode ser feito publicitando o produto em causa às pessoas certas, pessoas que, por exemplo, conheçam o maior número de outras com as características desejadas.

Os conceitos de clique e de relaxações de cliques, relevantes na análise de redes sociais, e que modelam grupos fortemente coesos e de interesse, são encontrados na literatura, de uma forma dispersa. O trabalho desta dissertação foi motivado com o intuito de agregar, organizar e relacionar alguns destes conceitos levando à elaboração de um texto de apoio a este módulo da disciplina. Os capítulos 2 e 3 desta tese, resultam do texto referido, elaborado em conjunto com o Professor Doutor Luís Gouveia.

No entanto, o trabalho desta dissertação foi além desta motivação, contendo um capítulo extra, o capítulo 4, que estuda medidas de centralidade de nodos e de grafos. Este capítulo tem como base documentos da literatura, mas apresenta algumas extensões a esses documentos e conclusões distintas das que estes apresentam.

A dissertação reúne as mais usuais relaxações de cliques e medidas de centralidade. Em ambos os temas são apresentados diversos resultados conhecidos da literatura. No entanto, e como se mencionou antes, este trabalho é uma compilação de conceitos e resultados não tendo sido possível, no que diz respeito a alguns resultados, encontrar as provas originais. Em alguns casos, as provas desses resultados foram refeitas. Pretende-se que as provas apresentadas sejam curtas, simples e recorram a raciocínios semelhantes, de forma a criar uma noção de unidade. Pretende-se também, com esta intenção, que as provas sejam feitas com base apenas na informação contida na dissertação, pelo que mesmo as provas que estão presentes na literatura são recriadas no texto para cumprir estas condições.

Palavras-chave: Grafo, clique, k-plex, k-clique, k-club, medidas de centralidade

Abstract

The graphs module of the discipline "Introduction to Mathematical Models" studies the application of graphs to the analysis of social networks. In this perspective, the nodes of a graph represent people and an edge between two nodes indicates that the corresponding people know each other. Analyzing groups of individuals and the relationships between them can be particularly useful when these individuals are "undesirable", as is the case with criminals or terrorists. It can also be useful in building marketing strategies that aim to reach people with certain characteristics. This process can be done by advertising the product in question to the right people, people who, for example, know the greatest number of others with the desired characteristics.

The concepts of clique and clique relaxations, relevant in the analysis of social networks, and which model strongly cohesive and groups of interest, are found in the literature, in a dispersed way. The work of this dissertation was motivated in order to aggregate, organize and relate some of these concepts leading to the development of a text to support this module of the discipline. Chapters 1 and 2 of this dissertation are the result of the aforementioned text, prepared jointly with Professor Doutor Luís Gouveia.

However, the work of this dissertation went beyond this motivation, containing an extra chapter, chapter 4, which studies measures of centrality of nodes and graphs. This chapter is based on literature documents but presents some extensions to these documents and conclusions different from those they present.

The dissertation brings together the most usual clique relaxations and centrality measures. In both themes, several results known from the literature are presented. However, and as mentioned before, this work is a compilation of concepts and results and it was not possible, with respect to some results, to find the original proof. In some cases, the proof of these results has been redone. It is intended that the proofs presented are short, simple and use similar lines of thought, to create a sense of unity. It is also intended, that the proofs are made based only on the information contained in the dissertation, reason why even the tests that are present in the literature are recreated in the text to fulfill these conditions.

Keywords: Graph, clique, k-plex, k-clique, k-club, centrality measure

Índice

1. Introdução.....	1
1.1. Estado da arte sobre redes sociais	1
1.2. Revisão bibliográfica de cliques e relaxações de cliques e medidas de centralidade.....	2
1.3. Organização da dissertação	3
2. Revisão de conceitos e definição de notação	5
3. Cliques e relaxações de cliques	11
3.1. Caracterização polinomial de maximalidade.....	11
3.2. Clique	12
3.2.1. Maximalidade de uma clique	13
3.3. Relaxações de Cliques.....	13
3.3.1. Relaxar o número de arestas.....	14
3.3.1.1. μ -quasi-clique.....	14
3.3.1.2. Maximalidade de uma μ -quasi-clique	15
3.3.2. Relaxar o grau dos nodos	16
3.3.2.1. k -plex.....	16
3.3.2.2. Maximalidade de um k -plex	17
3.3.3. Relaxar a distância entre nodos	17
3.3.3.1. k -clique.....	17
3.3.3.2. Maximalidade de uma k -clique	18
3.3.3.3. k -club.....	19
3.3.4. Relações entre conjuntos	21
3.3.4.1. Relação entre k -plex e μ -quasi-clique	21
3.3.4.2. Relação entre k -plex e k -club	23
3.3.4.3. Relação entre k -club e k -clique	26
3.4. Sumário	26
4. Medidas de centralidade.....	27
4.1. Medidas de centralidade de nodos.....	27
4.1.1. Centralidade de grau.....	27
4.1.2. Centralidade de proximidade.....	28
4.1.3. Centralidade de eficiência	28
4.1.3.1. Medida absoluta e relativa:.....	28
4.1.4. Centralidade de intermediação	29

4.1.5. Relações entre medidas de centralidade de nodos.....	30
4.2. Medidas de centralidade de grafos	31
4.2.1. Limites superiores de medidas de centralidade absolutas de grafos.....	32
4.2.1.1. Centralidade de grau.....	32
4.2.1.2. Centralidade de proximidade.....	33
4.2.1.3. Centralidade de eficiência	35
4.2.1.4. Centralidade de intermediação	36
4.3. Sumário	36
5. Conclusões	37
5.1. Síntese conclusiva	37
5.2. Possíveis desenvolvimentos futuros.....	37
6. Bibliografia.....	39

Índice de Figuras

Figura 2.1: Grafo G1	5
Figura 3.1: Grafo G2	13
Figura 3.2: Grafo G3	14
Figura 3.3: Grafo G4	16
Figura 3.4: Grafo G5	18
Figura 3.5: Grafo G6	19
Figura 3.6: Grafo G5	20
Figura 3.7: Grafo G7	20
Figura 3.8: Grafo G8	22
Figura 3.9: Grafo G9	25
Figura 3.10: Grafo G10	25
Figura 4.1: Grafo G11 (exemplo de [26])	30
Figura 4.2: Grafo G12	30
Figura 4.3: Medidas de centralidade	31

Tabela de notação

Notação	Objeto	Referência
$G = (V, A)$	Grafo	Definição 2.1
$grau(x)$	Grau do nodo x	Definição 2.4
$grau_S(x)$	Grau do nodo x no conjunto de nodos S	Definição 2.4
$dist(x, y)$	Distância entre os nodos x e y	Definição 2.6
$dist_S(x, y)$	Distância entre os nodos x e y no conjunto de nodos S	Definição 2.6
$ecc(i)$	Excentricidade do nodo i	Definição 2.7
$R(G)$	Raio do grafo G	Definição 2.8
$D(G)$	Diâmetro do grafo G	Definição 2.9

1. Introdução

Na presente dissertação são estudadas propriedades de conjuntos de nodos inseridos num grafo, extrapolando dos resultados desse estudo conclusões para o âmbito das redes sociais. Assim sendo, são feitos ao longo do texto paralelismos com este contexto. Deste modo, os nodos de um grafo representam indivíduos e uma aresta entre dois nodos representa uma relação de conhecimento entre os indivíduos respetivos.

1.1. Estado da arte sobre redes sociais

Nos dias de hoje as redes sociais desempenham um papel relevante na vida dos cidadãos, existindo redes sociais de cariz profissional e recreativo. Atualmente, devido aos diferentes tipos de uso associados, cada indivíduo poderá deter vários perfis numa panóplia de redes sociais distintas.

Sendo a interação em redes sociais algo tão comum, esta atividade gera informação que pode ser útil para empresas que pretendem vender um produto, para entidades de segurança que pretendam investigar as intenções ou ações de um dado indivíduo, para sociólogos que pretendam estudar relações entre indivíduos, etc.

As primeiras redes sociais na web terão surgido na década de 90. Pensa-se que tenha sido a rede social “ClassMates.com” a primeira a surgir. Esta página tinha como objetivo ajudar a reencontros de antigos colegas de escola/faculdade. Desde então muitas outras redes sociais foram desenvolvidas e grandemente utilizadas. Six Degrees surge em 1997, Friendster em 2002, MySpace e LinkedIn em 2003, e desde então e ainda muito populares atualmente, surgem o Facebook, o Twitter, Tumblr, Instagram, Google+, e muitas outras.

Apesar do aparecimento das redes sociais na web ter acontecido apenas na década de 90, há muito que são estudadas redes sociais. Os primeiros estudos de redes sociais remetem para as décadas de 30 e 40, estudos de sociometria por Jacob Moreno em 1934 [1] e de “balanço” numa rede de relações por Fritz Heider em 1946 [2]. No trabalho de Moreno são apresentadas as relações entre pessoas por meio de sociogramas, que são na verdade, mas ainda não referidos como tal, grafos. Neste trabalho discutem-se formas de medir relações entre indivíduos, determinar grupos e subgrupos, etc. O objetivo desta análise seria o de quantificar e determinar a estrutura de relacionamentos num sistema social.

O balanço discutido por Heider tem que ver com a existência de sentimentos positivos e negativos entre um conjunto de indivíduos e a classificação desse grupo enquanto grupo equilibrado ou desequilibrado tem por base esses sentimentos. Heider influenciou os trabalhos de Harary, Cartwright e Norman, [3] e [4], nas décadas de 50 e 60, que já incorporam nos seus estudos fundamentos em teoria de grafos, unindo os sociogramas de Moreno e as ideias de Heider.

Em 1954, John Barnes [5], antropólogo, deixou também a sua marca na História da evolução da análise de redes sociais, utilizando pela primeira vez o termo “social network”. No seu trabalho, Barnes estuda redes sociais com o objetivo de compreender a influência que têm as relações entre indivíduos num sistema social. Nadel, em [6], continua o desenvolvimento das ideias de Barnes, frisando a importância da estrutura de uma rede social versus o seu conteúdo e determinando o papel de cada elemento num grupo por via das relações que detém. Nadel [6] sugere para esta análise ferramentas algébricas.

Mitchell, seguidor de Nadel, apresenta conceitos já relacionados com o que são hoje as medidas de centralidade em grafos. No seu trabalho [7], em [8], é feita uma reflexão sobre a densidade de relações no grupo e são descritas redes “egocêntricas” (ego-centred), redes que se ramificam em torno de um

indivíduo. Paralelamente o papel de um indivíduo numa rede social, após os trabalhos de Nadel, começou a ganhar importância na comunidade científica, despoletando a elaboração de estudos por parte de Harrison White, sociólogo, e alguns dos seus colegas. Estes trabalhos focaram-se em análises matemáticas de redes sociais, com recurso a teoria de grafos e álgebra de grupos e semi-grupos. White introduziu também modelos matemáticos de estruturas sociais, como por exemplo a modelação em blocos, [8] e [9].

O interesse de White pelo estudo de redes sociais influenciou colegas e alunos que mais tarde formaram um grupo internacional com o objetivo de reunir investigadores neste campo (International Network Society of Social Network Analysts). Foi então, a partir dos anos 70, que a área de análise de redes sociais se tornou mais comum e reconhecida na comunidade científica.

É na década de 90 que se verificam mudanças relevantes na área da análise de redes sociais. Neste período, concretamente em 1998, Watts e Strogatz's publicam um estudo de "Small-World Networks" [10]. Neste artigo são modeladas redes neuronais, redes de doenças infecciosas e muitas outras redes com recurso à análise de redes sociais. A análise de redes sociais propaga-se então para outras áreas científicas, tendo partido de estudos em psicologia, sociologia, e antropologia, passando pela matemática e reunindo agora características que revelam interesse para uma série de outras áreas, como física, biologia, economia, etc. Consequentemente, Otte e Rousseau, em 2002, revêm estudos de ciências médicas, sociológicas e físicas através de uma perspetiva de análise de redes sociais [11]. Desde então e até hoje a análise de redes sociais mantém-se como um tema de interesse na área de teoria de grafos e como uma ferramenta útil para estudos de outras ciências.

1.2. Revisão bibliográfica de cliques e relaxações de cliques e medidas de centralidade

O conceito de clique é introduzido na literatura por Luce e Perry em 49 [12] seguindo-se a introdução do conceito de k -clique, também por Luce, em 1950 [13]. Uma k -clique é um conjunto caracterizado tendo por base distâncias entre nodos, distâncias calculadas através de cadeias que podem englobar nodos exteriores ao conjunto. Para colmatar a falta de um conjunto que considerasse distâncias apenas com nodos do conjunto, surge o conceito de k -club, ainda com o nome de clique sociométrica de diâmetro k , apresentado por Alba em 1973 [14]. Este conceito é explorado de novo em 79 por Mokken [15]. Paralelamente, em 68, Barnes fala sobre a noção de densidade de arestas num grafo para analisar a familiaridade dentro de um grupo de indivíduos [16]. Esta densidade de arestas foi formalizada como μ -quasi-clique em 2002 por Abello, Resende e Sudarsky [17]. O conceito de k -plex surge num trabalho de Seidman e Foster em 78 [18].

A noção de centralidade é introduzida por Bavelas em 48 [19] e após a publicação de uma série de medidas de centralidade na literatura, Freeman, em 79, [20] resume estas medidas em apenas três, a centralidade de grau, centralidade de proximidade e centralidade de intermediação. A medida de centralidade de eficiência é introduzida mais tarde por Hage e Harary em 1995 [21].

Para os temas que compõem o presente trabalho, apesar de vastamente explorados na literatura, não se conhece uma compilação que englobe e relacione os principais conceitos a eles associados.

No entanto, no que diz respeito aos conteúdos do capítulo 0, um artigo de Balasundaram e Pajouh [22] cobre grande parte deles. Outros trabalhos que concernem também o tema do capítulo 0 são o artigo de Cavique, Mendes e Santos [23] onde se enunciam as definições de clique, k -clique, k -plex e k -club e o quinto capítulo do livro "Handbook of Optimization in Complex Networks" de Pattillo, Youssef, e Butenko [24] que define também um clique e relaxações de clique, não examinando, no entanto, as suas relações.

No que diz respeito aos conteúdos do capítulo 4, a maior referência é o documento [25].

A compilação apresentada, sob a forma de dissertação, é elaborada com o intuito de ser sucinta, simplificada, organizada e com notação e nomenclatura consistente. Ademais, engloba em paralelo uma interpretação no âmbito das redes sociais, em todo o texto, ajudando à compreensão dos conceitos e das relações entre eles.

1.3. Organização da dissertação

A presente dissertação é composta por 4 capítulos, sendo o primeiro a Introdução.

No segundo capítulo são apresentados diversos conceitos e notação utilizada no restante texto.

No terceiro capítulo são estudadas cliques e relaxações de cliques. Sendo um clique um conjunto de nodos todos eles adjacentes entre si, no contexto de redes sociais, representa um grupo de indivíduos todos eles relacionados. Este é o ponto de partida para a análise realizada no capítulo. Neste, são relaxadas algumas características que definem um clique, de modo a caracterizar conjuntos que a ela se assemelhem e que representem outros tipos de grupos de indivíduos. Deste modo, o terceiro capítulo divide-se em três secções. As duas primeiras dedicadas ao estudo da propriedade de clique e a terceira ao estudo dos conjuntos que derivam de relaxações desta propriedade.

As características de um clique relaxadas na terceira secção deste capítulo são o número de arestas, o grau dos seus nodos e a distância entre os nodos. No final da secção são analisadas relações entre os conjuntos apresentados, através do estudo de relações das propriedades que os definem. Paralelamente são expostas relações entre tipos de grupos de indivíduos.

No quarto capítulo são estudadas medidas de centralidade, numa primeira secção, medidas de centralidade de nodos e numa segunda, medidas de centralidade de grafos. São estudados vários tipos de centralidade: de grau, de proximidade, de eficiência e de intermediação.

As medidas de centralidade aplicadas aos nodos de um grafo definem, no contexto de redes sociais, medidas para determinadas características dos indivíduos. No caso da centralidade de grau, é medido o quão “popular” é um indivíduo, tendo por base o número de pessoas com quem tem relações de conhecimento. Através da centralidade de proximidade mede-se o quão rapidamente (através de muitos ou poucos contactos) um indivíduo consegue comunicar com os restantes elementos do grupo. A centralidade de eficiência mede o quão perto está um indivíduo do indivíduo que lhe está mais distante (em relações de conhecimento). A centralidade de intermediação mede o quanto um indivíduo exerce o papel de intermediador num grupo, se está presente num pequeno ou num grande número de cadeias de comunicação entre outros elementos.

À semelhança do que acontece com os conjuntos estudados no terceiro capítulo, no final da secção são estudadas relações entre as várias medidas de centralidade de nodos, com recurso a alguns exemplos.

Na segunda secção do terceiro capítulo as medidas de centralidade de nodos são adaptadas a grafos. Neste caso, medem o quão desequilibrado é um grafo nas características dos seus nodos. No contexto de redes sociais, o quanto cada uma das características expostas atrás está mais concentrada num só indivíduo (ou grupo de indivíduos) em detrimento dos restantes. Um grafo com alta centralidade de grau, por exemplo, é aquele em que um indivíduo tem muitas relações de conhecimento (um nodo tem um elevado grau) e em que os restantes têm poucas (baixo grau).

2. Revisão de conceitos e definição de notação

Para uma total compreensão do texto que se segue é necessário introduzir alguns conceitos nele utilizados. O grafo será a estrutura utilizada para representar uma rede social em todas as análises que se seguem, deste modo, define-se abaixo um grafo.

Definição 2.1: Um grafo é um par ordenado $G = (V, A)$ em que V é um conjunto finito e A é um subconjunto de $\{\{i, j\}: i, j \in V, i \neq j\}$. Os elementos de V designam-se por nodos do grafo. Os elementos de A designam-se por arestas do grafo.

Um exemplo da definição anterior é o grafo $G1 = (V, A)$ em que $V = (1,2,3,4,5,6)$ e $A = \{\{1,2\}, \{2,3\}, \{3,4\}, \{3,5\}, \{4,6\}, \{5,6\}\}$. Este é também o grafo que será utilizado para exemplificar os restantes conceitos apresentados no presente capítulo.

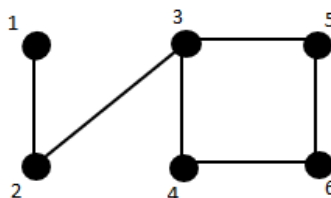


FIGURA 2.1: GRAFO G1

Na Figura 2.1 os nodos do grafo G1 são representados por pontos e as arestas por linhas. Esta é a representação adotada no documento.

Existem também grafos que, ao invés de arestas, possuem arcos. Os primeiros designam-se por grafos não orientados e os segundos por grafos orientados. As arestas definem-se como ligações entre pares de nodos i e j e representam-se $\{i, j\}$. Dois nodos que fazem parte de um destes pares dizem-se adjacentes e extremos da aresta definida pelo par. Os arcos são também pares de nodos, mas, neste caso, com uma ordem. Para dois nodos i e j os arcos (i, j) e (j, i) são dois objetos distintos. Em representações semelhantes à da Figura 2.1 estes objetos representam-se por setas. Neste texto não são abordados grafos orientados visto que, nos grafos estudados, os nodos representam pessoas e as arestas relações que conhecimento entre elas, que se consideram bilaterais.

Existem também grafos com arestas do tipo $\{i, i\}$, designadas por lacetes. Este tipo de arestas também não é incluído no presente texto. Isto porque, neste texto, tal como referido, as arestas representam relações de conhecimento entre indivíduos. Ora, não tem sentido a existência de uma relação de conhecimento de um indivíduo consigo próprio. Também não tem sentido, neste contexto, considerar múltiplas relações de conhecimento entre o mesmo par de indivíduos, pelo que entre dois nodos poderá apenas existir uma aresta. Grafos em que não existe mais de uma aresta com os mesmos extremos dizem-se grafos simples.

Os grafos utilizados no texto serão, de acordo com o estipulado, simples, não orientados e sem lacetes.

Uma definição que será ser interessante para a compreensão dos conceitos explorados neste texto é a de grafo completo. Apresenta-se esta definição abaixo.

Definição 2.2: Um grafo completo é um grafo $G = (V, A)$ simples e sem lacetes em que todos os nodos são adjacentes, ou, equivalentemente, um grafo simples e sem lacetes em que $|A| = \binom{|V|}{2}$.

Um grafo completo representa uma rede social em que todos os indivíduos se conhecem (todos os nodos são adjacentes).

Na presença de uma rede social poder-se-á querer analisar as relações de um determinado grupo de pessoas, ao invés das relações entre todas as pessoas que a compõem. Deste modo é importante saber-se extrair de um grafo, um outro grafo, de menor dimensão, que contenha apenas os nodos que representam pessoas de interesse e as arestas que representem as relações entre elas. Introduce-se abaixo o conceito de subgrafo induzido por um conjunto de nodos, motivado por esta necessidade.

Definição 2.3: Seja $G = (V, A)$. G' é subgrafo de G induzido por um conjunto de nodos $S \subset V$ se $G' = (S, A(S))$ onde $A(S) = \{\{x, y\} \in A : x, y \in S\}$.

A designação $A(S)$ para o conjunto de arestas com ambos os extremos no conjunto de nodos S continuará a ser utilizada no texto, sem nova menção do seu significado.

No grafo G_1 (Figura 2.1) pode considerar-se o subgrafo $G_1' = (S, A(S))$ induzido pelo subconjunto de nodos $S = \{1, 3, 5, 6\}$ e em que $A(S) = \{\{3, 5\}, \{5, 6\}\}$.

Um subgrafo é também um grafo pelo que pode ser completo, representando um grupo de pessoas, numa rede social, em todas elas se conhecem.

Para estudar uma rede social, um dado que pode ser importante considerar é o número de indivíduos que cada indivíduo da rede conhece. Num grafo este número corresponde à quantidade de arestas com um dos extremos no nodo que representa o indivíduo. Esta quantidade designa-se por grau do nodo.

Poder-se-á também querer considerar a quantidade de indivíduos que um indivíduo conhece num grupo restrito. Olhando ao conceito de subgrafo induzido por um conjunto de nodos, é intuitivo que essa quantidade corresponda ao grau do nodo que representa o indivíduo num subgrafo. Este grau, dentro de um subconjunto de nodos S , designa-se por grau do nodo em S .

Abaixo definem-se os conceitos de grau de um nodo (num grafo) e grau de um nodo num subconjunto de nodos (no subgrafo induzido por este conjunto).

Definição 2.4: Seja $G = (V, A)$ e $S \subset V$. O grau de um nodo $x \in V$ define-se como o número de arestas incidentes nesse nodo e representa-se por $grau(x)$, sendo que uma aresta é incidente num nodo se esse nodo é um dos seus extremos. Para qualquer nodo $x \in S$, denote-se por $grau_S(x)$ o número de arestas incidentes em x e com o outro extremo também em S , ou seja, o número de arestas incidentes em x que pertencem a $A(S)$.

No Grafo G_1 , e considerando o conjunto $S = \{1, 3, 5, 6\}$, $grau(3) = 3$ e $grau_S(3) = 1$.

Algo que também poderá ser importante analisar numa rede social é o comprimento do “caminho” que uma mensagem percorre de um emissor a um recetor. Isto é, se um indivíduo de um grupo quiser fazer chegar uma mensagem a um outro, recorrendo ao menor número possível de contactos

(seja por uma questão de rapidez, segurança ou integridade da mensagem) quantos contactos são necessários?

Para que se possa responder a esta pergunta é necessário definir primeiramente o conceito de cadeia.

Definição 2.5 Designa-se por cadeia num grafo $G = (V, A)$, uma sequência alternada de nodos e arestas $i_1, a_1, i_2, a_2, \dots, a_{k-1}, i_k$ tal que para $i = 1, \dots, k - 1$, se tem que $a_i = (i_i, i_{i+1})$. Os nodos i_1 e i_k designam-se por extremos da cadeia. Designa-se por ciclo uma cadeia em que os nodos i_1 e i_k coincidam.

A dimensão da menor cadeia que une dois nodos denomina-se distância. A distância diz respeito à soma do valor das arestas que compõem a cadeia. Neste caso, sendo uma aresta um contacto, considera-se que o seu valor é uma unidade. Mais uma vez, a distância entre nodos, tal como o grau de um nodo, é um conceito que pode ser aplicado a apenas parte de um grafo, a um subgrafo induzido por um conjunto de nodos. Neste caso a distância entre dois nodos corresponde ao número mínimo de contactos por onde passa uma mensagem entre os indivíduos representados por esses nodos, mas, desta vez, recorrendo apenas a contactos entre indivíduos do grupo representado pelo subgrafo.

Caso não exista possibilidade de transmitir uma mensagem de um indivíduo para outro, considera-se que a distância entre os nodos que os representam é infinita.

Os conceitos de distância (num grafo) e de distância num conjunto de nodos (no subgrafo por ele induzido) definem-se abaixo.

Definição 2.6: Seja $G = (V, A)$ e $x, y \in V$. A distância entre dois nodos, x e y , $dist(x, y)$, define-se como o comprimento da cadeia mais curta entre estes nodos. Caso não exista nenhuma cadeia entre x e y , assume-se $dist(x, y) = \infty$. Caso x coincida com y a distância é nula, isto é, $\forall x \in V$, $dist(x, x) = 0$. Seja $S \subset V$. Denota-se por $dist_S(x, y)$, e designa-se por distância entre x e y em S , o comprimento da cadeia mais curta entre x e y que inclui apenas nodos em S . A distância entre um nodo $w \notin S$ e o conjunto de nodos S define-se como $dist(w, S) = \min_{i \in S} \{dist(w, i)\}$.

Note-se que, com k escalar, $dist_S(x, y) \leq k \Rightarrow dist(x, y) \leq k$, visto que o comprimento da cadeia mais curta entre x e y que inclui apenas nodos de S é de certo igual ou superior ao da cadeia mais curta entre x e y que não contempla esta restrição. No contexto das redes sociais, tal significa que o número mínimo de contactos necessário para fazer chegar uma mensagem, de um indivíduo a outro, usando apenas contactos entre indivíduos de um grupo restrito, nunca poderá ser menor do que no caso de essa restrição não ser considerada.

No grafo G1 (Figura 2.1), apresentam-se abaixo exemplos de distâncias e distâncias num subconjunto de nodos.

Seja $S = \{1, 3, 5, 6\}$. Sendo a cadeia mais curta entre 1 e 3, a cadeia $\{\{1, 2\}, \{2, 3\}\}$, $dist(1, 3) = 2$. No entanto, $dist_S(1, 3) = \infty$, visto não existirem cadeias entre 1 e 3 que incluam apenas nodos em S . Confirma-se $dist(1, 3) = 2 \leq d_S(1, 3) = \infty$.

Uma questão que se pode levantar partindo do conceito de distância é a seguinte: de entre todas as distâncias a que um nodo está dos restantes, qual é a maior? Ou seja, partindo uma mensagem de um indivíduo, e sabendo que esta percorrerá o menor caminho possível para chegar ao destinatário, qual o

maior caminho que poderá ter de percorrer? Este comprimento de caminho, ou distância, associada a um nodo (indivíduo) designa-se por excentricidade do nodo.

Se todos os contactos diretos entre indivíduos numa rede social se efetuarem com uma rapidez semelhante, a excentricidade de um nodo indica o quão rapidamente o indivíduo por ele representado consegue comunicar com aquele que lhe está mais longínquo (cujo alcance requer mais contactos intermédios).

Um menor número de contactos na transmissão de uma mensagem pode também significar uma menor probabilidade de distorção da mesma. Assim, a excentricidade de um nodo pode indicar o nível de preservação de uma mensagem enviada pelo respetivo indivíduo no pior caso (caso em que esteja a comunicar com o indivíduo que lhe está mais longínquo).

Define-se abaixo a noção de excentricidade. As medidas anteriores, grau e distância, serão analisadas tanto em todo o grafo como num seu subconjunto de nodos, simultaneamente, no capítulo 0. A medida de excentricidade será apenas utilizada no capítulo 4 deste texto, onde, por razões de simplificação, se analisará sempre o grafo representativo do grupo de pessoas em estudo, seja este um subgrafo de um outro grafo, ou não.

Definição 2.7: Seja $G = (V, A)$. A excentricidade de um nodo $i \in V$ denota-se $ecc(i)$ e corresponde à maior distância a que o nodo i está de qualquer nodo de V . Isto é, $ecc(i) = \max_{j \in V} \{dist(i, j)\}$.

No grafo G1, calcule-se a excentricidade do nodo 2.

$$\begin{aligned} ecc(2) &= \max_{j \in V} \{dist(i, j)\} \\ &= \max\{dist(2,1), dist(2,2), dist(2,3), dist(2,4), dist(2,5), dist(2,6)\} \\ &= \max\{1,0,1,2,2,3\} = 3 \end{aligned}$$

Sabendo o quão distante está cada indivíduo daquele que lhe for mais longínquo, isto é, sabendo a excentricidade de cada um dos nodos de um grafo, poder-se-á querer saber qual a menor dessas distâncias/ excentricidades. O indivíduo associado ao nodo que tiver a menor excentricidade será aquele cujo pior caso, em termos de distância aos restantes, é o melhor de todos.

Se, por exemplo, se pretender seleccionar um indivíduo num grupo para que seja o ponto de origem de todas as mensagens, e se, tal como mencionado anteriormente, a rapidez ou a integridade das mensagens forem fatores importantes, então, uma boa escolha será a do indivíduo representado pelo nodo com menor excentricidade. O valor da excentricidade desse nodo, num grafo, designa-se por raio do grafo, conceito definido abaixo.

Definição 2.8: O raio de um grafo G é a menor das excentricidades dos seus nodos. Ou seja, $R(G) = \min_{i \in V} \{ecc(i)\}$.

Calcule-se o raio do Grafo G1:

$$\begin{aligned} R(G1) &= \min_{i \in V} \{ecc(i)\} = \min\{ecc(1), ecc(2), ecc(3), ecc(4), ecc(5), ecc(6)\} \\ &= \min \{4,3,2,3,3,4\} = 2 \end{aligned}$$

Outra questão que se pode levantar é, ao invés de qual a menor excentricidade num grafo, qual a maior. Este valor corresponderá ao comprimento do maior caminho que uma mensagem poderá ter de percorrer, independentemente de qual a sua origem ou destino. De novo, este comprimento, poderá indicar qual o maior intervalo de tempo tomado pela transmissão de uma mensagem, no pior caso, ou,

também no pior caso, o grau de risco de degradação de uma mensagem. A maior das excentricidades dos nodos de um grafo designa-se por diâmetro do grafo, conceito definido abaixo.

Definição 2.9: O diâmetro de um grafo G é a maior das excentricidades dos seus nodos. Ou seja, $D(G) = \max_{i \in V} \{ecc(i)\}$.

Calcule-se o diâmetro do Grafo G1:

$$\begin{aligned} D(G1) &= \max_{i \in V} \{ecc(i)\} = \max\{ecc(1), ecc(2), ecc(3), ecc(4), ecc(5), ecc(6)\} \\ &= \max\{4, 3, 2, 3, 3, 4\} = 4 \end{aligned}$$

3. Cliques e relaxações de cliques

Neste capítulo abordam-se subconjuntos de nodos com interesse no estudo de redes sociais. Estes subconjuntos de nodos num grafo representam grupos de pessoas numa rede social. Cada um dos subconjuntos enunciados define-se à custa de uma determinada propriedade, uma característica cujo significado no contexto de rede social será sempre mencionado e analisado.

Num grafo, vários subconjuntos de nodos podem verificar uma determinada propriedade, podendo estes ter dimensões diferentes e podendo alguns deles estar contidos noutros. Duas noções distintas se podem associar a estes conjuntos. Poder-se-á averiguar se cada um destes conjuntos é máximo e se cada um destes conjuntos é maximal. A primeira noção, a de conjunto máximo, tem a ver com a cardinalidade do conjunto. Um conjunto máximo para uma dada propriedade é aquele que tem maior número de nodos de entre todos os conjuntos que verificam a propriedade. Por sua vez, um conjunto maximal para uma dada propriedade é um conjunto que não está contido em nenhum outro que verifique essa propriedade. Abaixo apresentam-se, seguidos, os conceitos de conjunto máximo e de conjunto maximal.

Definição 3.1: Seja $G = (V, A)$. Um conjunto de nodos $S \subset V$ é conjunto máximo para uma propriedade P se S é o conjunto de maior cardinalidade, em V , que satisfaz essa propriedade.

Definição 3.2: Seja $G = (V, A)$. Um conjunto de nodos $S \subset V$ é conjunto maximal para uma propriedade P se não existe nenhum conjunto $M \subseteq V$ tal que $S \subset M$ e M verifica a propriedade P .

Ora, se um subconjunto de nodos for conjunto máximo para uma determinada propriedade, então, não existe nenhum outro de maior dimensão que verifique essa propriedade, pelo que não poderá existir nenhum outro que verifique a propriedade e que o contenha. Assim, um conjunto máximo é necessariamente maximal. O contrário não se verifica, o facto de um conjunto não estar contido em nenhum outro não significa que este conjunto seja máximo. Poderá existir um outro que não o contenha, que verifique a propriedade, e cuja cardinalidade seja maior.

No presente texto dar-se-á relevo à noção de maximalidade. Aquando da apresentação de cada uma das propriedades estudadas serão dados exemplos de subconjuntos máximos e maximais que as verificam. A noção de conjunto máximo restringir-se-á apenas a estes exemplos. A noção de maximalidade será aprofundada.

3.1. Caracterização polinomial de maximalidade

Tanto a determinação de conjuntos máximos como a de conjuntos maximais é um processo complexo, visto que, geralmente, não é possível fazê-lo por via de um algoritmo polinomial. Isto é, tanto uma como outra implicam que se verifique a satisfação da propriedade em causa por quase todos os subconjuntos de nodos do grafo em estudo. Claro que, por exemplo, na determinação de conjuntos máximos, pode fazer-se esta averiguação por ordem não crescente da cardinalidade dos conjuntos e, assim que seja determinado um conjunto que verifique a propriedade, não será necessário analisar os de cardinalidade menor. Diferentes técnicas e ordens de análise poderão ser aplicadas a esta tarefa, diminuindo a quantidade de conjuntos a analisar, no entanto, é sempre uma tarefa complexa e muito dependente daquilo que se vai descobrindo sobre o grafo ao longo do processo e de características específicas do mesmo.

Como referido, no presente trabalho, de entre estes dois conceitos, será o de maximalidade o mais focado. Isto porque, para algumas propriedades, existe um algoritmo para determinar conjuntos maximais. Para algumas propriedades, se um conjunto que verifica a propriedade não está contido em nenhum outro que tenha apenas mais um nodo e também a verifique, então não está contido em nenhum outro que a verifique, sendo, portanto maximal. Isto significa que partindo de um qualquer conjunto que verifique a propriedade, poder-se-á aumentá-lo nodo a nodo até que este seja maximal. Esta característica, que será designada como caracterização polinomial de maximalidade, também implica que ao remover um nodo de um conjunto que verifique a propriedade, este mantém a propriedade. Uma analogia que se pode fazer, de modo a simplificar a noção de caracterização polinomial de maximalidade é a de que um conjunto que verifique uma propriedade que goza desta caracterização, pode ser construído ou desconstruído como um lego, peça a peça (nodo a nodo), sem perda da propriedade. Abaixo define-se a caracterização polinomial de maximalidade.

Definição 3.3: Seja $G = (V, A)$. A propriedade P goza da caracterização polinomial de maximalidade se para $\forall S \subset V$ que satisfaça a P e seja maximal, $\forall y \in V \setminus S, S \cup \{y\}$ não verifica a propriedade P .

Considere-se um grafo $G = (V, A)$. Tal como referido, verificar a maximalidade de um conjunto $S \subset V$ que satisfaça uma propriedade P que esteja nas condições da caracterização da Definição 3.3, pode ser feito por inspeção, analisando no máximo $|V| - |S|$ conjuntos. Além disso, esta caracterização mostra uma forma eficiente de aumentar um conjunto não maximal até que este o passe a ser. Para tal, inclui-se em S um nodo y pertencente a $V \setminus S$ tal que $S \cup \{y\}$ verifica P e considera-se $S = S \cup \{y\}$. De seguida repete-se o processo, e assim sucessivamente, até que não exista nenhum nodo y em $V \setminus S$ tal que $S \cup \{y\}$ verifique a propriedade P .

Seja um conjunto S que satisfaça a propriedade P que não está nas condições da caracterização da Definição 3.3. Verificar a sua maximalidade por inspeção consiste em analisar, no pior caso, $\sum_{k=1}^{|V|-|S|} \binom{|V|-|S|}{k}$ conjuntos relativamente à propriedade P , ou seja, analisar todos os conjuntos que pertencem a $V \setminus S$, averiguando se a união de algum deles com S verifica ainda a propriedade P .

Ao longo do texto, e após a introdução de uma determinada propriedade, indicar-se-á em que situação se encontra um conjunto que satisfaça essa propriedade no que diz respeito à caracterização polinomial de maximalidade.

3.2. Clique

O primeiro conjunto que se apresenta é a clique. Este conjunto surge no contexto de redes sociais como indicador de grupos coesos de pessoas. Define-se abaixo uma clique.

Definição 3.4: Seja $G = (V, A)$. Uma clique define-se como um subconjunto de nodos vazio ou que induz um subgrafo completo. Assim, $S \subset V$ é clique sse $S = \{\}$ ou $\forall x, y \in S, \exists \{x, y\} \in A$.

Uma clique indica um conjunto de pessoas em que todas se relacionam umas com as outras. Ou seja, um conjunto de pessoas com interesses comuns. Se os nodos do grafo em questão representarem pessoas “indesejáveis”, uma clique pode ser indicadora de um conjunto de indivíduos que necessitam de ser vigiados, já que interações entre eles podem ser um indicador de situações de perigo. Abaixo exemplifica-se o conceito de clique com recurso a um novo grafo, o grafo G_2 .

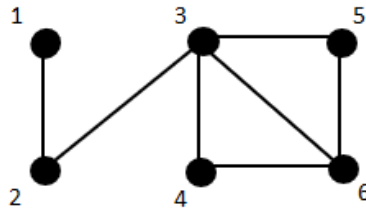


FIGURA 3.1: GRAFO G2

Em G_2 , o conjunto de nodos $\{1\}$ é uma clique não maximal, visto estar contida numa outra, a clique $\{1,2\}$. A clique $\{1,2\}$ é maximal porque não existe nenhuma outra que a contenha. No entanto, não é clique máxima, pois existem outras cliques, nomeadamente a clique $\{3,4,6\}$, com um maior número de nodos. Visto não existir nenhuma clique em G_2 com mais de 3 nodos, $\{3,4,6\}$ é clique máxima.

De seguida, tal como estipulado no início do capítulo, será averiguado se a propriedade de clique goza da caracterização polinomial de maximalidade.

3.2.1. Maximalidade de uma clique

Resultado 3.1: A propriedade de clique está nas condições da caracterização polinomial de maximalidade. Assim sendo, e assumindo $G = (V, A)$, uma clique $S \subset V$ é maximal sse $\forall y \in V \setminus S$, $S \cup \{y\}$ não é clique.

Prova:

Seja S uma clique. Remova-se um dos nodos de S , seja y , e designe-se o conjunto de nodos resultante por S' . A remoção do nodo y de S não influencia as ligações entre os restantes nodos do conjunto, mantendo-se os nodos de S' todos adjacentes entre si, pelo que S' é também uma clique. Continuando este raciocínio conclui-se que se S é clique então qualquer subconjunto de S é clique. Assim, se existe uma clique M que contenha S e $|M| - |S| \geq 2$, então, existe um y tal que $S \cup \{y\}$ é clique, o que prova que esta propriedade está nas condições da caracterização polinomial de maximalidade.

c.q.d.

Os subconjuntos de nodos que se seguem são incluídos num subcapítulo designado por “Relaxações de Cliques” visto terem origem na relaxação de várias propriedades que definem uma clique.

3.3. Relaxações de Cliques

Uma clique pode ser considerada uma estrutura muito restrita nos contextos indicados. Por exemplo, um grupo de 5 pessoas que contém um par de pessoas que não se conhecem, continua a ser um grupo coeso, ou indicador de um sinal de perigo, conforme a interpretação considerada. Além disso, em certas situações, ligações existentes entre pessoas são ainda desconhecidas (por exemplo, na base de dados de uma agência de informações). Estes casos levam à definição de subconjuntos de nodos que podem ser vistos como relaxações de cliques e que podem também ser considerados estruturas de interesse.

Seja $G = (V, A)$ um grafo e seja S uma clique.

Propriedades da clique S :

- i) $|A(S)| = \binom{|S|}{2} = \frac{|S|!}{2(|S|-2)!} = \frac{|S|(|S|-1)}{2}$
- ii) $\forall x \in S, grau_S(x) = |S| - 1$
- iii) $\forall x, y \in S, d_S(x, y) \leq 1$

Relaxar-se-á cada uma das propriedades enunciadas com vista a definir novas propriedades que caracterizem outros subconjuntos. A primeira propriedade a ser relaxada é a propriedade i), a que define o número de arestas existentes numa clique.

3.3.1. Relaxar o número de arestas

A primeira relaxação resulta imediatamente da motivação anterior. E se não existirem todas as ligações possíveis em S ? Poderá existir apenas uma percentagem dessa quantidade. Um subconjunto de nodos que contenha pelo menos $(\mu \times 100)\%$ da quantidade de arestas de uma clique com a mesma cardinalidade designa-se μ -quasi-clique.

3.3.1.1. μ -quasi-clique

Abaixo define-se o conceito de μ -quasi-clique.

Definição 3.5: Seja $G = (V, A)$. Uma μ -quasi-clique define-se como um subconjunto de nodos $S \subset V$ com pelo menos $\mu \binom{|S|}{2}$ arestas, sendo μ uma constante entre 0 e 1. Assim sendo, $S \subset V$ é μ -quasi-clique sse $|A(S)| \geq \mu \binom{|S|}{2}$.

Repare-se que uma μ -quasi-clique em que μ tem valor 1 é uma clique, pois a propriedade que a define é igual à propriedade i) que caracteriza uma clique.

Abaixo exemplifica-se o conceito de μ -quasi-clique com recurso ao grafo G_3 .

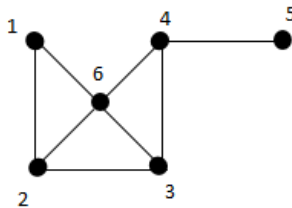


FIGURA 3.2: GRAFO G_3

Em G_3 considere-se o conjunto de nodos $S = \{1,2,3,4\}$.

Ora, $|A(S)| = |\{1,2\}, \{2,3\}, \{3,4\}| = 3$ e $\binom{|S|}{2} = \binom{4}{2} = 6$, logo, S é uma 0.5 –quasi-clique, no entanto, não é maximal, pois está contida em $S \cup \{6\}$ e em $S \cup \{5,6\}$, que são também 0.5 –quasi-cliques. Visto não existir em G_3 nenhum conjunto de nodos com maior cardinalidade do que $S \cup \{5,6\}$, este conjunto é a 0.5 –quasi-clique máxima.

Veja-se, de seguida, como se comportam os conjuntos que gozam da propriedade de μ -quasi-clique no que diz respeito à caracterização polinomial de maximalidade.

3.3.1.2. Maximalidade de uma μ -quasi-clique

Resultado 3.2: A propriedade de μ -quasi-clique está nas condições da caracterização polinomial de maximalidade. Assim sendo, e assumindo $G = (V, A)$, $S \subset V$ é μ -quasi-clique maximal sse $\forall y \in V \setminus S$, $S \cup \{y\}$ não é μ -quasi-clique.

Prova:

Considere-se $S \subset V$ uma μ -quasi-clique. Suponha-se que $\forall y \in V \setminus S$, $S \cup \{y\}$ não é μ -quasi-clique, e veja-se se $S \cup M$, tal que $M \subset V \setminus S$, $y \in M$ e $|M| \geq 2$, poderá ser μ -quasi-clique. Seja $|M| = m$.

Ora,

- i) sendo S uma μ -quasi-clique, $|A(S)| \geq \mu \binom{|S|}{2}$;
- ii) não sendo $S \cup \{y\}$ uma μ -quasi-clique, $|A(S \cup \{y\})| < \mu \binom{|S|+1}{2}$

Veja-se, então, se é possível que $|A(S \cup M)| \geq \mu \binom{|S|+m}{2}$.

Pode obter-se o número de arestas do subgrafo induzido por $S \cup M$ somando o número de arestas dos subgrafos induzidos por $S \cup \{y_i\}$, tal que $y_i \in M$, e subtraindo $m - 1$ vezes o número de arestas em S (visto que estas são somadas em todos os conjuntos $S \cup \{y_i\}$).

Assim,

$$|A(S \cup M)| = \sum_{i=1}^m |A(S \cup \{y_i\})| - (m - 1) |A(S)|$$

Como

$$|A(S \cup \{y_i\})| < \mu \binom{|S| + 1}{2}, \forall y_i \in V \setminus S, \text{ e } |A(S)| \geq \mu \binom{|S|}{2}$$

então,

$$\begin{aligned} |A(S \cup M)| &= \sum_{i=1}^m |A(S \cup \{y_i\})| - (m - 1) |A(S)| < \sum_{i=1}^m \mu \binom{|S| + 1}{2} - (m - 1) \mu \binom{|S|}{2} \\ &= m \mu \binom{|S| + 1}{2} - (m - 1) \mu \binom{|S|}{2} = m \mu \frac{(|S| + 1)|S|}{2} - (m - 1) \mu \frac{|S|(|S| - 1)}{2} \\ &= \mu \frac{m(|S| + 1)|S| - (m - 1)|S|(|S| - 1)}{2} = \mu \frac{m|S|^2 + m|S| - (m - 1)(|S|^2 - |S|)}{2} \\ &= \mu \frac{m|S|^2 + m|S| - (m|S|^2 - m|S| - |S|^2 + |S|)}{2} \\ &= \mu \frac{m|S|^2 + m|S| - m|S|^2 + m|S| + |S|^2 - |S|}{2} = \mu \frac{2m|S| + |S|^2 - |S|}{2} \\ &= \mu \frac{|S|(2m - 1 + |S|)}{2} \end{aligned}$$

Veja-se se $|S|(2m - 1 + |S|) < (|S| + m)(|S| + m - 1)$.

$$\begin{aligned}
& |S|(2m - 1 + |S|) < (|S| + m)(|S| + m - 1) \Leftrightarrow \\
& \Leftrightarrow 2m|S| - |S| + |S|^2 < |S|^2 + m|S| - |S| + m|S| + m^2 - m \Leftrightarrow \\
& 0 < m^2 - m
\end{aligned}$$

Chega-se a uma proposição verdadeira, visto que $m \geq 2$.

Conclui-se, portanto, que

$$|A(S \cup M)| < \mu \frac{|S|(2m - 1 + |S|)}{2} < \frac{(|S| + m)(|S| + m - 1)}{2} = \mu \binom{|S| + m}{2}$$

Assim, não sendo $S \cup \{y\}$ uma μ -quasi-clique, $\forall y \in V \setminus S$, então, $S \cup M$, $\forall M \subseteq V \setminus S$, não é μ -quasi-clique, pelo que μ -quasi-clique está nas condições da caracterização polinomial de maximalidade.

Isto é, S é μ -quasi-clique maximal sse $\forall y \in V \setminus S$, $S \cup \{y\}$ não é μ -quasi-clique.

c.q.d

3.3.2. Relaxar o grau dos nodos

Nesta secção considera-se um subconjunto de nodos que resulta de eliminar arestas de uma clique, mas em que existe um limite no número de arestas que se podem remover de um qualquer nodo. Assim, a propriedade a ser relaxada nesta secção é a propriedade ii) que caracteriza uma clique.

3.3.2.1. k -plex

Abaixo define-se o conceito de k -plex.

Definição 3.6: Seja $G = (V, A)$. Um k -plex define-se como um subconjunto de nodos $S \subset V$ em que, para cada nodo em S , há no máximo k nodos de S a que este não é adjacente. Assim sendo, $S \subset V$ é k -plex sse $\forall x \in S$, $grau_S(x) \geq |S| - k$, com $1 \leq k \leq |S|$.

Um k -plex representa um grupo de pessoas em que cada uma delas conhece, pelo menos, todas as outras exceto $k - 1$. A condição de grau para uma clique, $\forall x \in S$, $grau_S(x) = |S| - 1$, pode ser reescrita como $grau_S(x) \geq |S| - 1$, visto que $grau_S(x) \leq |S| - 1$, pelo que um 1-plex é uma clique.

A constante k tem de ser necessariamente igual ou superior a 1, visto que o grau máximo que um nodo pode ter num grafo (simples) é $|S| - 1$. Considera-se também que $k \leq |S|$ porque o grau de cada nodo é sempre maior ou igual a zero.

Atente-se no grafo G4.

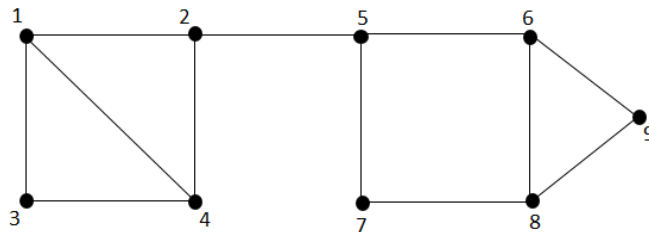


FIGURA 3.3: GRAFO G4

Em G_4 , o conjunto de nodos $\{1,2,3,4\}$ é um 2-plex pois todos os nodos que o compõem têm grau, dentro do conjunto, superior ou igual a $|S| - 2 = 4 - 2 = 2$. O mesmo se passa com o conjunto de nodos $\{5,6,7,8\}$.

Para que um conjunto com 5 nodos seja um 2-plex é necessário que todos os nodos que o compõem tenham, dentro do conjunto, grau igual ou superior a 3. Da mesma forma, para que um conjunto com 6 nodos seja 2-plex é necessário que todos os nodos que o compõem tenham, dentro do conjunto, grau 4 ou superior (o raciocínio repete-se para conjuntos com um maior número de nodos).

Nenhum nodo do grafo G_4 tem grau maior do que 3 pelo que não poderão existir 2-plex's com mais de 5 nodos. Como não se observa nenhum conjunto com 5 nodos e grau, dentro do conjunto, superior ou igual a 3, conclui-se que o 2-plex com maior cardinalidade tem 4 nodos. Assim sendo, $\{1,2,3,4\}$ e $\{5,6,7,8\}$ são 2-plex's máximos, e, consequentemente, maximais.

Averigüe-se de seguida se os conjuntos que gozam da propriedade de k -plex estão nas condições da caracterização polinomial de maximalidade.

3.3.2.2. Maximalidade de um k –plex

Resultado 3.3: A propriedade de k -plex está nas condições da caracterização polinomial de maximalidade. Assim sendo, e assumindo $G = (V, A)$, $S \subset V$ é k -plex maximal sse $\forall y \in V \setminus S$, $S \cup \{y\}$ não é k -plex.

Prova

Seja S um k -plex. Remova-se um dos nodos de S , diga-se y , e designe-se o conjunto de nodos resultante por S' . Ora, no máximo, a remoção do nodo y de S leva a uma redução em uma unidade do grau dos restantes nodos do conjunto. Assim, $\forall x \in S'$, $grau_{S'}(x) \geq |S| - k - 1 = |S'| - k$, pelo que S' é também k -plex. Continuando este raciocínio conclui-se que se S é k -plex então qualquer subconjunto de S é k -plex, o que prova que esta propriedade está nas condições da caracterização polinomial de maximalidade.

c.q.d.

3.3.3. Relaxar a distância entre nodos

Nesta secção consideram-se subconjuntos de nodos que resultam de eliminar arestas de uma clique, mas em que existe um limite no número de arestas que se podem remover em cadeias entre dois nodos. Deste modo define-se um limite máximo de distância entre quaisquer dois nodos.

Podem definir-se dois tipos de conjuntos através da imposição de um limite na distância entre nodos. Tal como mencionado no segundo capítulo do texto, quando em análise estão nodos que pertencem a um subconjunto, podem considerar-se dois tipos de distâncias diferentes. Pode considerar-se a distância entre os dois nodos dentro desse subconjunto ou em todo o grafo. A imposição de um limite máximo no primeiro tipo de distância é uma condição mais restrita do que essa imposição no segundo tipo de distância. Por este motivo, o primeiro subconjunto de nodos apresentado toma por base a segunda distância mencionada, a distância entre nodos em todo o grafo.

3.3.3.1. k –clique

Defina-se o conceito de k -clique.

Definição 3.7: Seja $G = (V, A)$. Uma k -clique define-se como um subconjunto de nodos $S \subset V$ em que, para quaisquer dois nodos $x, y \in S$, existe uma cadeia que os une, em G , com no máximo k arestas. Assim sendo, $S \subset V$ é k -clique sse $\forall x, y \in S, dist(x, y) \leq k$, com $1 \leq k \leq |V| - 1$.

Uma k -clique representa um conjunto de pessoas em que cada pessoa necessita de no máximo k contactos (arestas) para comunicar com outra.

Note-se que a análise da propriedade de k -clique não tem sentido num grafo desconexo, visto que neste existem nodos cuja distância entre eles tem valor infinito. Assim sendo, k , a maior distância entre dois nodos distintos de uma k -clique, é finita e varia entre 1 e $|V| - 1$.

É de notar também que uma 1-clique é uma clique.

Observe-se o grafo $G_5 = (V, A)$.

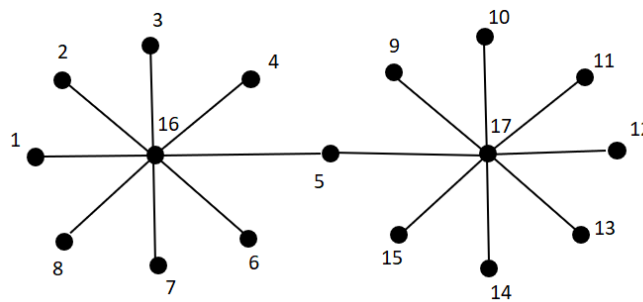


FIGURA 3.4: GRAFO G_5

O conjunto de nodos $\{1,2,3,4\}$ é uma 2-clique pois todos os nodos do conjunto estão, em G_5 , à distância de, no máximo 2 arestas. V é uma 4-clique pois todos os nodos de V estão, no máximo, à distância de 4 arestas.

Analise-se, à semelhança do que é feito com os restantes conjuntos, se a propriedade de k -clique está nas condições da caracterização polinomial de maximalidade.

3.3.3.2. Maximalidade de uma k -clique

Resultado 3.4: A propriedade de k -clique está nas condições da caracterização polinomial de maximalidade. Assim sendo, e assumindo $G = (V, A)$, $S \subset V$ é k -clique maximal sse $\forall y \in V \setminus S, S \cup \{y\}$ não é k -clique.

Prova

Considere-se uma k -clique S , tal que $S \subset V$. Suponha-se que existe um conjunto $Q \subset V \setminus S$, com $|Q| \geq 2$, e tal que $S \cup Q$ ainda é k -clique. Suponha-se também que $\forall y \in Q \setminus S, S \cup \{y\}$ não é k -clique.

Se $S \cup \{y\}$ não é k -clique então $\exists x \in S$ tal que $dist(x, y) > k$. Assim, $Q \cup S$ não é k -clique porque x e y também pertencem a esse conjunto, o que contradiz o ponto de partida.

Conclui-se que a propriedade de k -clique está nas condições da caracterização polinomial de maximalidade. **c.q.d.**

Feito o estudo da maximalidade de k -cliques, apresenta-se o segundo subconjunto caracterizado por uma relaxação da propriedade iii), e o último subconjunto de nodos, estudado neste texto, que é

relaxação de uma clique. Este subconjunto resulta de relaxar, não a distância entre os nodos no grafo, mas a distância entre os nodos no subconjunto onde estão inseridos, o subconjunto a classificar.

Numa k -clique os contactos referidos na sua descrição, que são no máximo k arestas, podem ter extremos que não estão em S . Quer isto dizer que apesar de se tratar de um grupo em que todas as pessoas conseguem comunicar-se por via de outras, estas outras podem não pertencer ao grupo, o que o fragiliza. A segurança no envio de mensagens entre elementos do grupo pode ser comprometida quando feito por meio de indivíduos externos ao grupo.

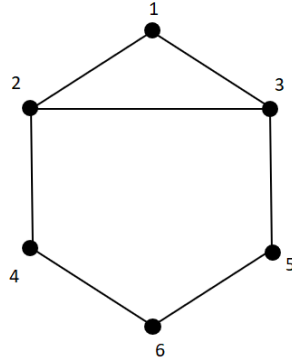


FIGURA 3.5: GRAFO G6

No grafo G_6 , se considerarmos a 2-clique $S = \{1,2,3,4,5\}$ verificamos que apesar da distância entre quaisquer dois nodos do conjunto ser no máximo 2, o diâmetro do mesmo é 3 pois, os nodos 4 e 5 distam 2 arestas no grafo G_6 e 3 em S .

Se considerarmos a 2-clique $S = \{2,3,6\}$ verificamos que o subgrafo induzido por este conjunto de nodos é desconexo. Os indivíduos associados aos nodos 2 e 3 só contactam 6 por via de outros indivíduos exteriores ao grupo. Surge assim a necessidade de uma estrutura em que as distâncias consideradas entre dois nodos sejam as distâncias que incluem só nodos do conjunto. Esta estrutura tem o nome de k -club e é uma relaxação de clique em que existe também um limite de arestas que se podem considerar em cadeias entre dois nodos, mas neste caso considera-se apenas as cadeias que incluem nodos do conjunto.

3.3.3.3. k – club

Define-se abaixo o conceito de k -club.

Definição 3.8: Seja $G = (V, A)$. Um k -club define-se como um subconjunto de nodos $S \subset V$ em que, para quaisquer dois nodos $x, y \in S$ existe uma cadeia que os une, em S , com no máximo k arestas. Assim sendo, $S \subset V$ é k -club sse $\forall x, y \in S, dist_S(x, y) \leq k$, com $1 \leq k \leq |S| - 1$.

A análise da propriedade de k -club num conjunto de nodos cujo subgrafo por eles induzido não é conexo não tem sentido pois neste existirão nodos cuja distância entre eles tem valor infinito. Assim a propriedade de k -club é definida apenas para conjuntos de nodos cujas distâncias entre eles (definidas por cadeias que incluem apenas nodos do conjunto) são finitas. Deste modo k é uma constante e varia entre 1 e $|S| - 1$, não sendo nunca negativa, por definição do conceito de distância. k é uma constante também sempre menor que $|S|$, visto que a distância de dois nodos num grafo conexo é no máximo $|S| - 1$.

É de notar também que um 1 – club é uma clique.

Recorde-se o grafo G5.

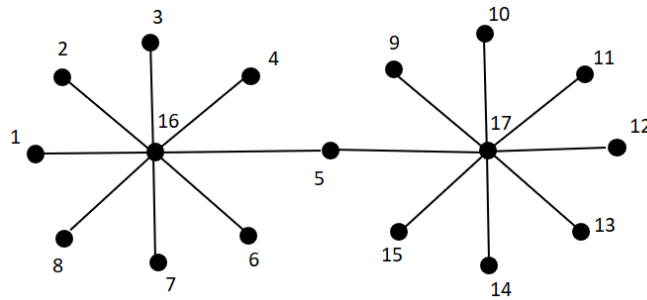


FIGURA 3.6: GRAFO G5

Note-se que o conjunto $M = \{1,2,3,4\}$, identificado anteriormente como uma 2-clique não é um 2-club. As cadeias de dimensão 2 que ligam nodos de M têm como nodo intermédio o nodo 16 que não pertence a M . Note-se também que não existem cadeias entre os nodos de M que não incluam nodos externos a M . O conjunto $\{1,2,3,4,16\}$ já é um 2-club.

Veja-se se a propriedade de k -club está nas condições da caracterização polinomial de maximalidade. Para tal atente-se no grafo G7.

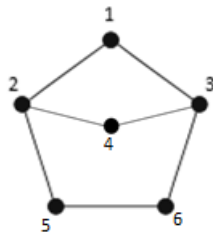


FIGURA 3.7: GRAFO G7

No grafo G7 considere-se os 2-clubs $S = \{1,2,3,4\}$ e $M = \{1,2,3,4,5,6\}$.

$S \subset M$, logo S não é maximal. No entanto, $S \cup \{5\}$ e $S \cup \{6\}$ não são 2-clubs. Percebe-se então que neste caso não poderá ser utilizado o raciocínio das k -cliques. Tal acontece porque as cadeias consideradas para averiguar se o nodo exterior ao conjunto verifica a propriedade diferem consoante os nodos que já estão incluídos no conjunto. Neste caso o nodo 5 não poderia ser adicionado a S porque em $S \cup \{5\}$ as cadeias que ligam 5 a 3 têm 3 arestas. Com o nodo 6 passa-se algo análogo. No entanto se ambos forem adicionados, proporcionam um ao outro, cadeias com menos arestas que permitem fazer a ligação entre eles e os restantes nodos do conjunto. Estando o nodo 6 incluído no conjunto, poder-se-á chegar do nodo 5 ao nodo 3 passando pelo 6 e, portanto, percorrendo só 2 arestas.

Conclui-se que, se um k -club $S \subset V$, tal que $G = (V, A)$, for maximal, então, $\forall x \in V \setminus S \exists y \in S$ tal que $dist_S(x, y) > k$, no entanto esta condição não é suficiente para que S seja maximal.

Resultado 3.5: A propriedade de k -club não está nas condições da caracterização polinomial de maximalidade.

Apresentados todos os subconjuntos (relaxações de clique) a considerar, interessa estudar as relações que existem entre as propriedades que os definem.

3.3.4. Relações entre conjuntos

Nesta subsecção analisam-se relações entre os conjuntos referidos, é estudada a existência de relações de implicação entre a satisfação das diversas propriedades enunciadas. Esta análise surge como resposta à questão “Visto que os conjuntos que satisfazem estas propriedades são todos conjuntos tão mais unidos quanto menor for o “ k ”, não obrigará a satisfação da condição de um a satisfazer a condição de outro?”

Comece-se pela relação entre os dois primeiros tipos de conjuntos apresentados, a μ -quasi-clique e o k -plex.

3.3.4.1. Relação entre k -plex e μ -quasi-clique

Apresenta-se abaixo um resultado que relaciona as propriedades de k -plex e de μ -quasi-clique.

Resultado 3.6: Assumindo $G = (V, A)$, $S \subset V$ e $|S| \geq 2$, se S é k -plex, com $k \geq 1$, então, S é μ -quasi-clique com $\mu \leq \frac{|S|-k}{|S|-1}$.

Prova

Considere-se um k -plex S e veja-se se o número de arestas de S o define como $\frac{|S|-k}{|S|-1}$ -quasi-clique.

Note-se que se pode calcular o número de arestas num conjunto de nodos S , $|A(S)|$, somando os graus dos nodos nesse conjunto e dividindo por 2. Tal verifica-se porque ao somar o grau de cada nodo se soma duas vezes cada aresta no conjunto, esta é contabilizada na soma dos graus dos dois nodos que une. Assim,

$$|A(S)| = \sum_{x \in S} \frac{\text{grau}_S(x)}{2}$$

Sabendo que $\text{grau}_S(x) \geq |S| - k$, $\forall x \in S$, então,

$$|A(S)| \geq \frac{\sum_{x \in S} (|S| - k)}{2} = \frac{|S|(|S| - k)}{2}$$

Para que S seja μ -quasi-clique, S tem que de se verificar

$$|A(S)| \geq \mu \binom{|S|}{2} = \mu \frac{|S|(|S| - 1)}{2}, \quad 0 \leq \mu \leq 1$$

Tem-se que

$$|A(S)| \geq |S| \frac{|S| - k}{2}$$

Para que S seja μ -quasi-clique, ter-se-á de verificar

$$\frac{|S|(|S| - k)}{2} \geq \mu \frac{|S|(|S| - 1)}{2} \Leftrightarrow \frac{|S| - k}{|S| - 1} \geq \mu$$

Conclui-se que se S é k -plex, com $k \geq 2$, então, S é μ -quasi-clique com $\mu \leq \frac{|S|-k}{|S|-1}$.

c.q.d.

Algo que é interessante perceber neste resultado é qualidade do valor $\frac{|S|-k}{|S|-1}$.

Será a satisfação da propriedade de k -plex por parte de um conjunto S condição suficiente para que este conjunto seja uma μ -quasi-clique com $\mu > \frac{|S|-k}{|S|-1}$, ou será $\frac{|S|-k}{|S|-1}$ o maior valor que se pode garantir para μ ?

Uma forma de verificar que $\frac{|S|-k}{|S|-1}$ é o maior valor que se pode garantir, é ver que $\frac{|S|-k}{|S|-1} \times \frac{|S|(|S|-1)}{2} = \frac{|S|(|S|-k)}{2}$ arestas é o suficiente para que S seja k -plex, é este o número de arestas em S quando cada nodo tem grau $|S| - k$.

Note-se que o recíproco do Resultado 3.6 não se verifica. Um bom contraexemplo para provar esta questão é grafo G_8 .

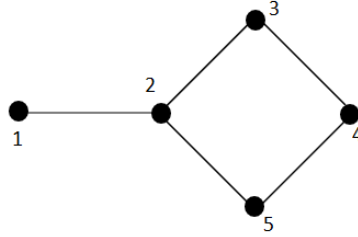


FIGURA 3.8: GRAFO G_8

Considere-se em G_8 o conjunto de nodos $S = \{1,2,3,4,5\}$. Seja $k = 3$.

S é $\frac{1}{2}$ -quasi-clique, onde $\frac{1}{2} = \frac{|S|-k}{|S|-1}$. No entanto, S não é 3-plex. Veja-se, por exemplo, que $\text{grau}(1) = 1 < |S| - 3 = 5 - 3 = 2$. Tal acontece porque as arestas de G_8 estão mais concentradas no conjunto de nodos $\{2,3,4,5\}$, deixando o nodo 1 com grau baixo. Um conjunto Q que é uma $\frac{|Q|-k}{|Q|-1}$ -quasi-clique será um k -plex se $\forall x, y \in Q, \text{grau}(x) = \text{grau}(y)$.

O resultado apresentado e demonstrado anteriormente diz algo muito simples e intuitivo no contexto das redes sociais. O Resultado 3.6 diz que, se dentro de um grupo cada uma das pessoas conhece muitas outras então a quantidade de relações de conhecimento no grupo é grande. Mais concretamente, sabe-se que se um grupo tem dimensão n e cada indivíduo do grupo conhece pelo menos $n - k$ pessoas, então existem no grupo pelo menos $\frac{n(n-k)}{2}$ relações de conhecimento, o que representa $\left(\frac{n-k}{n-1} \times 100\right)\%$ da quantidade máxima possível de relações de conhecimento para esse grupo.

O contrário não se verifica, isto é, uma quantidade alta de relações de conhecimento dentro de um grupo não implica uma quantidade alta de relações de conhecimento por parte de cada indivíduo. Esta noção também é de fácil compreensão, basta ver que as relações de conhecimento podem estar mais concentradas numa parte do grupo do que noutra.

A próxima relação entre conjuntos a analisar é a relação entre k -plex e k -club.

3.3.4.2. Relação entre k -plex e k -club

Recorde-se que um k -plex com um k “pequeno” representa um grupo onde cada elemento conhece quase todos os outros elementos. Isso não significará que são precisos no máximo k contactos para chegar de um elemento a outro? Ora, para tal é necessário que seja possível a comunicação entre quaisquer dois elementos. Será esta também condição suficiente, isto é, para que um k -plex seja k -club basta que seja conexo? A resposta a esta questão é dada pelo seguinte.

Resultado 3.7: Assumindo $G = (V, A)$ e $S \subset V$, se S é k -plex conexo, então, S é k -club.

Prova

Seja $G = (V, A)$ um grafo e S um 2-plex conexo contido em V .

Sejam x, y e w nodos de S tal que $\text{grau}(x) = \text{grau}(y) = \text{grau}(w) = |S| - 2$.

Ora x é adjacente a todos os nodos de S menos 1, tal como y e w . Consideremos que y é esse nodo a que x não está ligado. x e y são então, ambos adjacentes a w . Então, para se chegar de x a y são precisas 2 arestas, $\{x, w\}$ e $\{w, y\}$. S é, portanto, também um 2-club.

Veja-se se esta relação se estende a k -plex's de k maior.

Seja $G = (V, A)$ um grafo e seja agora S um k -plex conexo com $k > 2$.

Sejam x e y nodos de S .

Para se chegar de x a y serão necessárias mais do que k arestas de $A(S)$?

Ora comece-se a “caminhar” de x e considere-se a cada passo o pior caso possível. Isto é, x não é adjacente a y , então, avance-se de x para um nodo a , a não é adjacente a y , avance-se para um nodo b . Repetindo este processo e chegados ao $(k - 1)$ ésimo nodo a contar de x (exclusivé), teremos percorrido $(k - 1)$ arestas, $\{x, a\}, \{a, b\}, \{b, c\}, \dots, \{h, i\}$. Este $(k - 1)$ ésimo nodo terá de ser adjacente a y pois y é adjacente a todos os nodos de S , menos k (ele próprio e $k - 1$ outros nodos). Estes $k - 1$ nodos serão então x, a, b, \dots, i .

Conclui-se, então, que um k -plex, conexo, é um k -club.

c.q.d.

No contexto de redes sociais o Resultado 3.7 exprime que, se não existirem num grupo pares de indivíduos que não se conseguem contactar, poder-se-á determinar o número máximo de contactos necessários à comunicação entre quaisquer dois indivíduos com base no número mínimo de relações de conhecimento que cada um deles tem. Isto é, dentro dum grupo com dimensão n , se o mínimo de pessoas que um qualquer indivíduo conhece é $n - k$, então no máximo, serão precisos k contactos para que uma mensagem seja transmitida entre quaisquer dois indivíduos.

Note-se que, para que um qualquer conjunto seja k -club, o subgrafo induzido por esse conjunto tem de ser conexo, visto que a propriedade de k -club está definida apenas para conjuntos deste tipo (em que as distâncias entre nodos são finitas). Um limite superior para k que assegura que um k -plex é

conexo é o valor $\left\lceil \frac{|S|}{2} \right\rceil$. Assim, um k -plex tal que $k \leq \left\lceil \frac{|S|}{2} \right\rceil$ é k -club. E, como se pode verificar pelo resultado seguinte, além de k -club, este conjunto é também 2-club.

Resultado 3.8: Assumindo $G = (V, A)$ e $S \subset V$, se S é k -plex tal que $k \leq \left\lceil \frac{|S|}{2} \right\rceil$, então, S é 2-club.

Prova

Seja $G = (V, A)$ um grafo, S um k -plex contido em V e x e y nodos de S .

Assuma-se que $\{x, y\} \notin A$. Considere-se o conjunto de adjacentes do nodo x e o conjunto de adjacentes do nodo y . Assuma-se que a interseção destes conjuntos é vazia. Visto que o grau de cada um destes nodos é no mínimo $|S| - k$, e que nem x nem y fazem parte de nenhum dos conjuntos, para que a interseção dos mesmos seja vazia, $|S|$ terá de ser igual ou superior a $(|S| - k) \times 2 + 2 \geq 2 \left(|S| - \left\lceil \frac{|S|}{2} \right\rceil \right) + 2$.

Caso $|S|$ seja par,

$$2 \left(|S| - \left\lceil \frac{|S|}{2} \right\rceil \right) + 2 = 2 \left(|S| - \frac{|S|}{2} \right) + 2 = |S| + 2 > |S|.$$

Caso $|S|$ seja ímpar,

$$2 \left(|S| - \left\lceil \frac{|S|}{2} \right\rceil \right) + 2 = 2 \left(|S| - \left(\frac{|S|}{2} + \frac{1}{2} \right) \right) + 2 = |S| - 1 + 2 = |S| + 1 > |S|.$$

Tal não se verifica. Assim sendo, pelo menos um nodo de S está na interseção dos adjacentes de x e de y , pelo que existe entre eles uma cadeia com 2 arestas.

Caso a aresta $\{x, y\}$ esteja em A , $dist(x, y) = 1$.

Conclui-se que quaisquer nodos de S estão à distância de, no máximo, 2 arestas, pelo que S é 2-club.

c.q.d.

Interprete-se o que significa o resultado anterior no contexto de redes sociais. Tal como se suspeitava, a existência de muitas relações de conhecimento num grupo, por parte de cada um dos indivíduos que o constitui, leva a que um qualquer contacto entre dois desses indivíduos possa ser feito sem recorrer a muitos intermediários. Mais concretamente, pelo Resultado 3.8, sabe-se que, se todos os indivíduos de um grupo conhecerem pelo menos metade das pessoas que o constituem, é preciso, no máximo, um intermediário na transmissão de uma qualquer mensagem.

Será que o contrário também se verifica? Será que um baixo número de contactos necessários à transmissão de uma mensagem dentro de um grupo implica que os indivíduos que o constituem conhecem grande parte dos outros? Por outras palavras, será que um k -club é um k -plex?

A resposta a esta questão é dada através da análise do grafo G_9 .

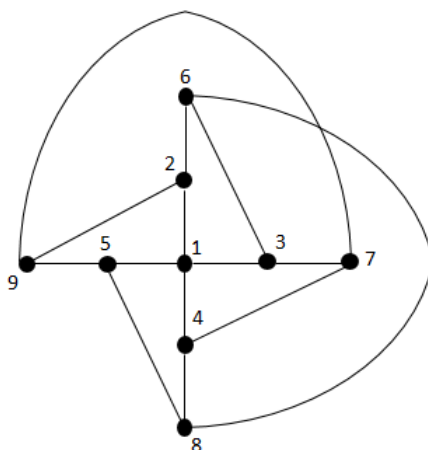


FIGURA 3.9: GRAFO G9

No Grafo G9, $V = \{1,2,3,4,5,6,7,8,9\}$ é um 2-club visto que todos os nodos estão à distância de no máximo 2 arestas. No entanto não é um 2-plex pois $grau(1) = 4 < |V| - 2 = 9 - 2 = 7$.

Depreende-se deste contraexemplo que um k -club nem sempre é um k -plex. Isto passar-se-á com todos os conjuntos de nodos que, tal como V , possuam um nodo adjacente a um pequeno grupo de nodos muito conectado com os restantes. Quer isto dizer que para que uma pessoa pertencente a um conjunto necessite de poucos contactos para chegar a todas as outras, esta não precisa de conhecer muitas outras, basta conhecer pessoas que conheçam muitas outras.

Poder-se-á descrever esta situação no contexto de grafos e de forma mais geral.

Seja $G = (V, A)$ um grafo conexo, sejam $x, w \in V$ e seja $k \in \mathbb{N}$.

Seja Q o conjunto de nodos a que x é adjacente, com $|Q| < |V| - k$, e tal que $w \notin Q$. Então $dist(w, x) = dist(w, Q) + dist(Q, x) \Leftrightarrow dist(w, x) = dist(w, Q) + 1$. Basta que a maior distância entre Q e qualquer nodo seja menor ou igual a $k - 1$ para que V seja um k -club. Assim, V é k -club e, como $grau(x) < |V| - k$, não é k -plex.

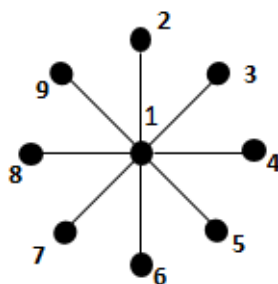


FIGURA 3.10: GRAFO G10

O conjunto de todos os nodos do grafo G10, um exemplo mais simples de um 2-club que não é 2-plex, é um caso particular do caso geral referido acima, tomando um nodo pertencente a $M = \{2,3,4,5,6,7,8,9\}$ como x e $Q = \{1\}$.

Considere-se $V = \{1,2,3,4,5,6,7,8,9\}$. Tome-se o nodo 2 como x .

Para que V seja 2-plex, $\forall y \in V, grau(y) \geq 9 - 2 = 7$. Visto que $grau(2) = |\{1\}| = 1 < 7$, V não é 2-plex. No entanto, a distância de um nodo w , não adjacente ao nodo 2, ao nodo 2, é de 2

unidades, $dist(w, 2) = dist(w, \{1\}) + dist(\{1\}, 2) \Leftrightarrow dist(w, 2) = dist(w, \{1\}) + 1$. Como $dist(w, \{1\}) = 1, \forall w \in V \setminus \{1, 2\}$, então, $dist(w, 2) = 2 \forall w \in V \setminus \{1, 2\}$. O mesmo se verifica se se considerar qualquer outro nodo de M como x . Assim, V é um 2-club.

Falta agora relacionar os conceitos de k -club e k -clique.

3.3.4.3. Relação entre k -club e k -clique

É fácil de notar que um k -club é uma k -clique e que o inverso não se verifica visto que, como mencionado no capítulo 2, “Revisão de conceitos e definição de notação”, o comprimento da cadeia mais curta entre x e y que inclui apenas nodos de S é igual ou superior ao da cadeia mais curta entre x e y que não contempla essa restrição. Assim, tem-se o seguinte resultado.

Resultado 3.9: Assumindo $G = (V, A)$ e $S \subset V$, se S é k -club, então, S é k -clique.

Na próxima subsecção resumem-se as relações estudadas entre os vários subconjuntos.

3.4. Sumário

Após o estudo dos conjuntos que são relaxações de cliques e das relações que existem entre eles, conclui-se que um k -plex é μ -quasi-clique com $\mu \leq \frac{|S|-k}{|S|-1}$. Conclui-se também que a propriedade de k -plex implica a propriedade k -club e, conseqüentemente, de k -clique, se o subgrafo induzido pelo conjunto em causa for conexo. Estas conclusões, no âmbito das redes sociais, ligam-se ao facto de um elevado número de relações por parte de cada indivíduo de um grupo implicar a existência de uma curta cadeia de contactos para comunicação entre cada dois destes indivíduos.

4. Medidas de centralidade

Se as definições anteriores estudavam a coesão de um grupo, as medidas de centralidade estudam a posição de indivíduos num grupo e o quanto este é equilibrado.

Primeiramente serão apresentadas algumas medidas de centralidade de nodos. Estas medidas pretendem responder a questões como: Quem será que dentro dum grupo é o elemento mais crucial no que diz respeito à comunicação entre outros? Quem será dentro dum grupo o elemento mais fácil de contactar, mais próximo de todos? Estas questões ganham particular importância quando se trata de um grupo de indivíduos perigoso, uma rede de terrorismo por exemplo.

Serão, então, apresentados vários tipos de centralidade que se podem medir num nodo e como pode ser feita essa medição. Para além de uma medida absoluta é sempre apresentada uma medida relativa, que permite perceber não só a diferença de centralidade entre nodos de um mesmo grafo, mas também a diferença de centralidade de nodos cujo valor absoluto de centralidade seja igual em grafos de dimensões diferentes. Quer isto dizer que a medida relativa informa, não só acerca do peso de um indivíduo relativamente a outros do grupo a que pertence, mas também do peso do indivíduo no grupo relativamente a um outro indivíduo num outro grupo de dimensão diferente.

As medidas apresentadas serão consideradas num grafo $G = (V, A)$ simples e conexo, sendo $|V| = n$.

4.1. Medidas de centralidade de nodos

4.1.1. Centralidade de grau

A centralidade de grau mede o quão “popular” é um indivíduo.

A medida corresponde ao grau do nodo. Neste caso um nodo é mais central quanto maior for o seu grau. Ou, por outras palavras, um indivíduo é mais central quanto mais pessoas conhecer.

Considere-se um grafo em que um indivíduo (nodo) detêm muitas relações de conhecimento (centralidade elevada) e os restantes não (centralidade baixa). Então, é provável que para que estes últimos se comuniquem tenham de usar o primeiro como intermediário. Note-se que esta situação é provável de acontecer num k -club que não é um k -plex, uma estrutura onde os nodos que a constituem não possuem muitas ligações a outros, mas em que estes se conseguem comunicar através de um número limitado de contactos.

Definição 4.1: A medida absoluta de centralidade de grau de um nodo $x \in V$ define-se como $C_D(x) = grau(x)$

Definição 4.2: A medida relativa de centralidade de grau de um nodo $x \in V$ define-se como $C_D^r(x) = \frac{grau(x)}{n-1}$

Para obter a medida relativa, o valor de centralidade de grau é dividido pelo valor máximo que pode tomar, sendo este $n - 1$, caso em que um nodo está ligado a todos os outros.

4.1.2. Centralidade de proximidade

A centralidade de proximidade mede o quão perto está um indivíduo de todos os outros. Ou, assumindo o tempo de contacto proporcional à distância, o quão rápida/facilmente consegue um indivíduo fazer chegar uma mensagem a todos os outros.

Esta medida corresponde ao inverso da soma das distâncias do nodo aos restantes. Assim, um nodo será tão mais central quanto menor for esta soma de distâncias.

No que diz respeito a esta medida de centralidade poder-se-á pensar também noutros contextos que não o de redes sociais. Por exemplo, a escolha de um local para construção de uma escola poderá passar pela análise da centralidade de proximidade de um conjunto de locais onde residem estudantes, visto ser um fator importante na localização de uma escola a distância a que esta se encontra dos seus alunos. Esta situação passa-se também com outros serviços públicos frequentes, como posto de correios etc.

Definição 4.3: A medida absoluta de centralidade de proximidade de um nodo $x \in V$ define-se como $C_C(x) = \frac{1}{\sum_{i=1}^n dist(i,x)}$.

Definição 4.4: A medida relativa de centralidade de proximidade de um nodo $x \in V$ define-se como $C_C^r(x) = \frac{n-1}{\sum_{i=1}^n dist(i,x)}$.

A medida relativa é obtida dividindo o valor de centralidade de proximidade pelo valor máximo que esta pode tomar, sendo este $\frac{1}{n-1}$, caso em que o nodo está à distância de uma aresta de todos os outros.

Note-se que a medida de centralidade de proximidade deixa de ser significativa se o grafo em análise for desconexo, pois haverá distâncias entre nodos com valor infinito. Assim, esta medida está definida apenas para grafos conexos.

4.1.3. Centralidade de eficiência

A centralidade de eficiência mede o quão perto está um indivíduo do indivíduo que lhe está mais distante.

A medida corresponde ao inverso da excentricidade do nodo, sendo este, portanto, tão mais central quanto menor for a sua excentricidade.

Esta medida, tal como a anterior, é interessante para contextos além de redes sociais. Se a medida de centralidade de proximidade pode ajudar à escolha de um local adequado para um serviço público frequente, a centralidade de eficiência pode ajudar com a escolha do local adequado para um serviço de urgências. É importante que um serviço de urgências não esteja longe de nenhum indivíduo, independentemente de estar próximo da maioria. Estes serviços poderão ser bombeiros, hospitais, etc.

4.1.3.1. Medida absoluta e relativa:

Definição 4.5: A medida absoluta de centralidade de eficiência de um nodo $x \in V$ designa-se por $C_E(x)$ e a medida relativa de centralidade de eficiência por $C_E^r(x)$. Ambas as medidas têm valor $\frac{1}{ecc(x)}$.

Neste caso a medida relativa corresponde à medida absoluta, visto que o valor máximo que a medida absoluta pode tomar é 1. A medida de centralidade de eficiência (absoluta) toma valor 1 quando o nodo que está mais longe do nodo em análise dista do mesmo apenas uma aresta. Ou seja, caso em que o nodo em análise é adjacente a todos os outros.

Note-se que, tal como a medida de centralidade de proximidade, a medida de centralidade de eficiência deixa de ser significativa se o grafo não for conexo, pelo que está definida somente para grafos conexos.

4.1.4. Centralidade de intermediação

A centralidade de intermediação mede o quanto um indivíduo exerce o papel de intermediador de informação num grupo.

A medida corresponde ao número de cadeias mais curtas que passam pelo nodo. Ou seja, ao número de vias de comunicação entre indivíduos que o indivíduo associado ao nodo pode intercetar, ou, a que consegue aceder. Esta medida mede mais claramente aquilo que é referido também na medida de centralidade de grau, a informação a que um indivíduo tem acesso. Note-se que para tirar estas conclusões se assume que as cadeias pelas quais viaja a informação são as mais curtas.

Para apresentar matematicamente a medida de centralidade de intermediação introduza-se primeiro a função de intermediação.

Defina-se g_{ij} como o número de cadeias mais curtas entre o nodo i e o nodo j e $g_{ij}(x)$ como o número de cadeias mais curtas entre i e j que incluem o nodo x .

Defina-se $b_{ij}(x)$ como a função de intermediação de i e j no ponto x tal que $b_{ij}(x) = \frac{g_{ij}(x)}{g_{ij}}$.

Definição 4.6 : A medida absoluta de centralidade de intermediação de um nodo $x \in V$ define-se como $C_B(x) = \sum_{1 \leq i < j \leq n, i, j \neq x} b_{ij}(x)$

Definição 4.7: A medida relativa de centralidade de intermediação de um nodo $x \in V$ define-se como $C_B^r(x) = \frac{2C_B(x)}{n^2 - 3n + 2}$

Tal como para as medidas anteriores, a medida relativa corresponde à absoluta sobre o seu valor máximo possível, neste caso, $\frac{n^2 - 3n + 2}{2}$.

A função de intermediação de i e j no ponto x , tal que $x \neq i, j$, é no máximo 1, no caso em que x pertence a todas as cadeias mais curtas entre os nodos i e j . Se x pertencer a todas as cadeias mais curtas entre todos os pares de nodos (que não o incluem), então $\sum_{1 \leq i < j \leq n, i, j \neq x} b_{ij}(x) = \sum_{1 \leq i < j \leq n, i, j \neq x} 1 = \left[\binom{n}{2} - (n - 1) \right] \times 1 = \frac{n^2 - n - 2n + 2}{2} = \frac{n^2 - 3n + 2}{2}$.

Conhecidas as várias medidas de centralidade dos nodos, na próxima subsecção estudam-se as relações entre elas.

4.1.5. Relações entre medidas de centralidade de nodos

Tal como com as propriedades de conjuntos estudadas no capítulo “Cliques e relaxações de cliques” medem a coesão desses conjuntos, as medidas de centralidade de nodos medem a relevância de um nodo numa rede. O que levanta a questão “Um nodo ter centralidade máxima, em G , de acordo com uma das medidas, significa que tem de acordo com outra medida?”

Esta pergunta deve-se a que é intuitivo pensar que um nodo com valor de grau elevado é um nodo que estará presente em muitas cadeias mais curtas e que está muito perto dos restantes, e vice-versa. Um grafo trivial nessas condições é um grafo com topologia em forma de estrela. No entanto, como se ilustra a seguir, a resposta à questão levantada acima é negativa, justificando, em certa medida, a existência e estudo, de medidas diferentes.

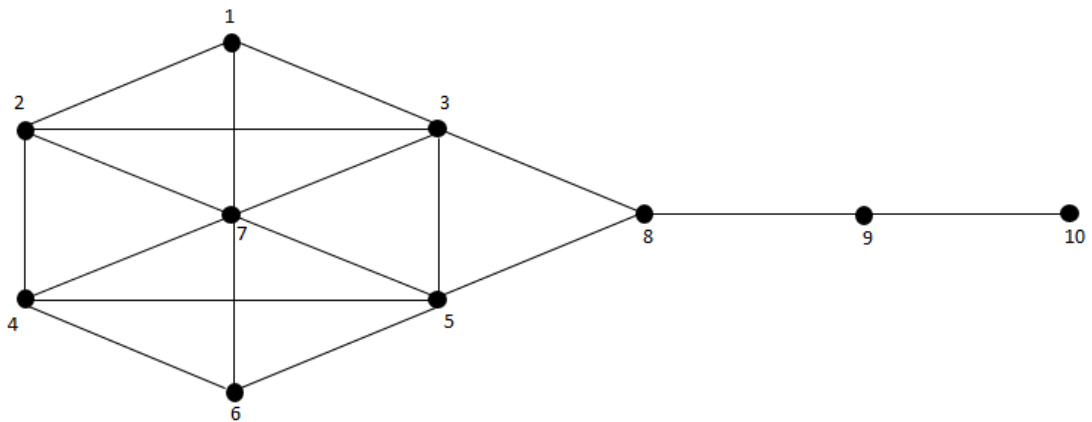


FIGURA 4.1: GRAFO G11 (EXEMPLO DE [26])

No grafo G11, pode observar-se que o nodo 7 é o que tem maior centralidade de grau, no entanto, devido à sua distância aos nodos 8,9 e 10 pode antecipar-se que não seja o nodo mais próximo de todos os outros. Calculando a centralidade de proximidade de todos os nodos conclui-se que 3 e 5 são os nodos que atingem o maior valor, ou seja, são estes os nodos atingíveis por todos os outros num número mínimo de passos (arestas). É também intuitivo que o nodo 8 seja o nodo com maior centralidade de intermediação, visto fazer a ponte entre os nodos 9 e 10 e os restantes, e tal, de facto, confirma-se. Neste caso o nodo 8 é também o nodo com maior eficiência, apesar das medidas de centralidade de eficiência e de intermediação não serem equivalentes, como se pode verificar no grafo abaixo.

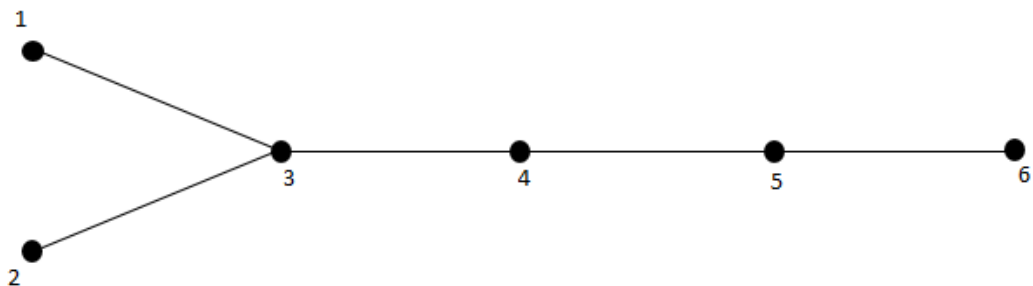
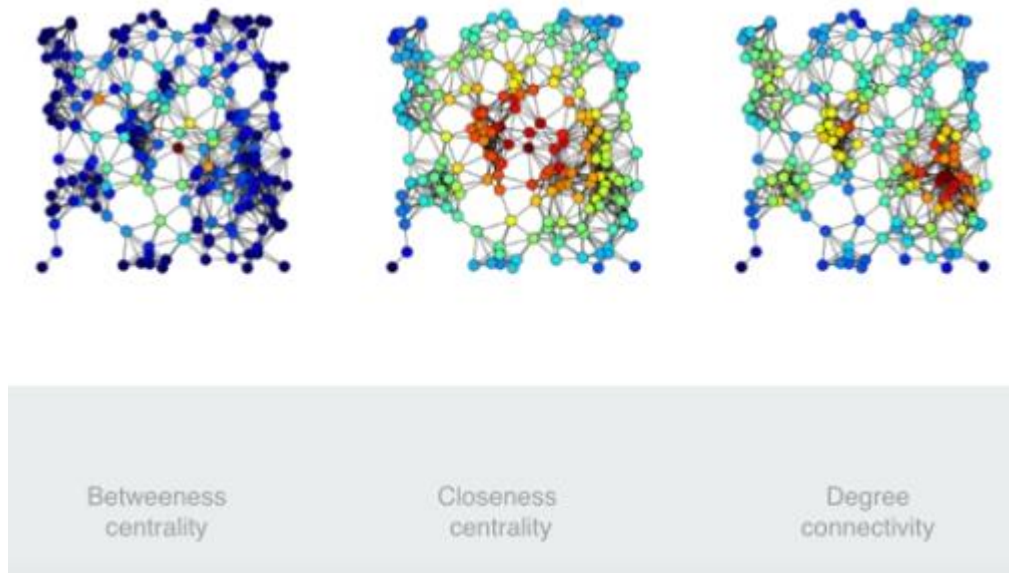


FIGURA 4.2: GRAFO G12

No grafo G12, o nodo 3 tem maior centralidade de intermediação e o nodo 4 tem maior centralidade de eficiência, visto que o primeiro pertence a um maior número de cadeias mais curtas e que a distância máxima entre o segundo e qualquer outro nodo é menor que a do primeiro.

Observe-se a diferença entre as medidas de centralidade de grau (degree), proximidade (closeness) e de intermediação (betweenness) num grafo de grande dimensão.



<https://www.youtube.com/watch?v=NgUj8DEH5Tc>

FIGURA 4.3: MEDIDAS DE CENTRALIDADE

Na Figura 4.3 a centralidade de cada nodo varia de baixa a elevada consoante a cor do nodo varie de azul escuro a vermelho. Observa-se que no grafo apresentado o conjunto de nodos que tem maior centralidade de proximidade difere do conjunto de nodos que tem maior centralidade de grau. Ademais, a centralidade de intermediação dos seus nodos é, no geral baixa, ao invés do que acontece com as duas outras medidas de centralidade.

Feito o estudo das medidas de centralidade dos nodos de um grafo parte-se, na próxima secção, para o estudo das medidas de centralidade de grafos.

4.2. Medidas de centralidade de grafos

Neste capítulo mede-se a centralidade de um grafo, ao invés da de um nodo. Ou seja, mede-se o quanto um grafo é equilibrado nas centralidades dos seus nodos, isto é, se existe um nodo ou grupo de nodos mais centrais do que os restantes ou se todos os nodos do grafo têm valores de centralidade semelhantes. Na primeira situação a centralidade do grafo é elevada, enquanto que na segunda é baixa.

Serão estudadas as medidas de centralidade de grau, proximidade, eficiência e intermediação.

Transpondo a informação anterior para uma rede social, uma rede representada por um grafo com alta centralidade é uma rede muito dependente de um indivíduo, ou conjunto de indivíduos.

Considere-se $G = (V, A)$ um grafo, defina-se o conceito de centro, e, seguidamente, o de centralidade de um grafo.

Definição 4.8: Seja $G = (V, A)$. O centro de G respeitante à medida de centralidade Y , representado por $\beta_Y(G)$, é o subconjunto de nodos com valor máximo da medida de centralidade Y em G . Ou seja, $\beta_Y(G) = \{x \in V: C_Y(x) \geq C_Y(z), \forall z \in V\}$.

Definição 4.9: Seja $G = (V, A)$ e seja $|V| = n$. Seja $x \in \beta_Y(G)$ e seja $C_Y(x)$ o valor da medida de centralidade Y do nodo x . Seja $\Phi_Y(G) = \sum_{i=1}^n (C_Y(x) - C_Y(i))$ e seja Ω o conjunto de todos os grafos com n vértices. Seja S_n o grafo estrela com n nodos. A centralidade de G respeitante à medida de centralidade Y , representada por $C_Y(G)$, define-se como o valor

$$\frac{\Phi_Y(G)}{\max\{\Phi_Y(H): H \in \Omega\}} = \frac{\Phi_Y(G)}{\Phi_Y(S_n)}$$

Analise-se a igualdade anterior.

$C_Y(G)$, como definido anteriormente, é a razão entre $\Phi_Y(G)$ e $\Phi_Y(H)$, sendo H o grafo que, tendo o mesmo número de nodos de G , maximiza a medida Φ_Y . Deste modo, poder-se-á ver a medida $C_Y(G)$ como uma medida relativa de centralidade de um grafo e $\Phi_Y(G)$ como a medida absoluta. Assim, tal como as medidas relativas de centralidade de nodos, $C_Y(G)$ varia entre 0 e 1 e informa do quão próximo está G de ter o valor máximo possível de centralidade atingível com o seu número de nodos.

Para um grafo G com n nodos, a igualdade anterior afirma que o grafo S_n tem centralidade absoluta ($\Phi_Y(S_n)$) máxima para qualquer uma das quatro medidas de centralidade Y . Porquê? Uma forma de responder a esta questão é a de encontrar, primeiramente, o valor máximo que Φ_Y pode tomar, para cada uma das medidas de centralidade apresentadas, e, de seguida, encontrar um grafo que satisfaça esse valor. Na subsecção seguinte é feito este processo.

4.2.1. Limites superiores de medidas de centralidade absolutas de grafos

Consideremos um grafo H com $n \geq 3$ nodos e estudemos as características que este grafo tem de ter de forma a maximizar $\Phi_Y(H)$ para cada medida de centralidade Y .

4.2.1.1. Centralidade de grau

Seja H um grafo com n nodos.

$$\Phi_D(H) = \sum_{i=1}^n (C_D(x) - C_D(i)).$$

Maximizando cada parcela, de forma independente, maximiza-se o somatório. Assim, torna-se claro que a situação ideal seria que um nodo tivesse centralidade máxima e todos os outros centralidade mínima. No que diz respeito à centralidade de grau, isto corresponde a que um nodo tenha centralidade $n - 1$ e os restantes, centralidade 1 (grau mínimo de um qualquer nodo num grafo conexo). Esta condição descreve um grafo estrela. Calcule-se, então, o valor de $\Phi_D(S_n)$

$$\begin{aligned}
\Phi_D(S_n) &= \sum_{i=1}^n (C_D(x) - C_D(i)) = \\
&= (C_D(x) - C_D(1)) + \dots + (C_D(x) - C_D(x)) + \dots + (C_D(x) - C_D(n)) \\
&= ((n-1) - 1) + \dots + ((n-1) - (n-1)) + \dots + ((n-1) - 1) \\
&= (n-1)(n-2) = n^2 - 2n - n + 2 = n^2 - 3n + 2
\end{aligned}$$

4.2.1.2. Centralidade de proximidade

Para encontrar o maior valor de $\Phi_C(H)$, ou seja, para maximizar a centralidade de proximidade de um grafo H , parta-se de novo de um grafo estrela com n nodos. Considere-se $H = S_n$, e averigue-se se é possível aumentar o valor da soma $\Phi_C(S_n)$ alterando a configuração do grafo.

Acrescentar novas arestas ao grafo resultará num aumento da centralidade de nodos não centrais o que apenas reduzirá a soma $\Phi_C(S_n)$.

Ao remover arestas do grafo estrela este ficará desconexo. No entanto, pode averiguar-se o efeito que tem no valor $\Phi_C(S_n)$ retirar uma aresta e introduzir uma outra que mantenha o grafo conexo. Seja x o nodo central, $\{x, y\}$ a aresta removida e $\{y, w\}$ a aresta adicionada ao grafo.

Seja $S_n = (V, A)$. Ora, esta alteração provocará o aumento de uma unidade na soma $\sum_{i=1}^n \text{dist}(i, x)$, o aumento em $n-3$ unidades na soma $\sum_{i=1}^n \text{dist}(i, y)$ e uma diminuição em uma unidade na soma $\sum_{i=1}^n \text{dist}(i, w)$. O aumento em $\sum_{i=1}^n \text{dist}(i, y)$ deve-se a que $n-3$ nodos do grafo deixam agora de estar a duas arestas de distância de y e passam a estar à distância de três arestas. Simultaneamente o nodo central, que lhe distava uma aresta passa a estar à distância de duas, e w , que estava à distância de duas arestas passa agora a estar à distância de uma aresta. Os restantes nodos ficarão agora à distância de três arestas do nodo y , ao invés de a duas como inicialmente. Assim, $\sum_{i=1}^n \text{dist}(i, j), \forall j \in V$, tal que $j \neq x, y, w$, aumentará uma unidade. Averigue-se o efeito destas alterações no somatório $\Phi_C(S_n)$.

Ao valor $\Phi_C(S_n)$ é subtraída $(n-1)$ vezes a quantidade $\frac{1}{\sum_{i=1}^n \text{dist}(i, x)} - \frac{1}{\sum_{i=1}^n \text{dist}(i, x) + 1}$, uma vez a quantidade $\frac{1}{\sum_{i=1}^n \text{dist}(i, w) - 1} - \frac{1}{\sum_{i=1}^n \text{dist}(i, w)}$ e somadas as quantidade $\frac{1}{\sum_{i=1}^n \text{dist}(i, y)} - \frac{1}{\sum_{i=1}^n \text{dist}(i, y) + n - 3}$ e $(n-3)$ vezes $\frac{1}{\sum_{i=1}^n \text{dist}(i, j)} - \frac{1}{\sum_{i=1}^n \text{dist}(i, j) + 1}$, sendo j um qualquer nodo de V , tal que $j \neq x, y, w$. Veja-se se o valor adicionado ao somatório $\Phi_C(S_n)$ é positivo.

$$\begin{aligned}
&-(n-1) \left(\frac{1}{\sum_{i=1}^n \text{dist}(i, x)} - \frac{1}{\sum_{i=1}^n \text{dist}(i, x) + 1} \right) - \left(\frac{1}{\sum_{i=1}^n \text{dist}(i, w) - 1} - \frac{1}{\sum_{i=1}^n \text{dist}(i, w)} \right) \\
&+ \frac{1}{\sum_{i=1}^n \text{dist}(i, y)} - \frac{1}{\sum_{i=1}^n \text{dist}(i, y) + n - 3} + (n-3) \left(\frac{1}{\sum_{i=1}^n \text{dist}(i, j)} - \frac{1}{\sum_{i=1}^n \text{dist}(i, j) + 1} \right) \\
&= -(n-1) \left(\frac{1}{n-1} - \frac{1}{n} \right) - \left(\frac{1}{2 \times (n-2) + 1 - 1} - \frac{1}{2 \times (n-2) + 1} \right) + \frac{1}{2 \times (n-2) + 1} \\
&\quad - \frac{1}{2 \times (n-2) + 1 + n - 3} + (n-3) \left(\frac{1}{2 \times (n-2) + 1} - \frac{1}{2 \times (n-2) + 1 + 1} \right) \\
&= -1 + \frac{n-1}{n} - \frac{1}{2n-4} + \frac{1}{2n-3} + \frac{1}{2n-3} - \frac{1}{3n-6} + \frac{n-3}{2n-3} - \frac{n-3}{2n-2}
\end{aligned}$$

$$\begin{aligned}
&= -\frac{1}{n} - \frac{1}{2n-4} + \frac{n-2}{2n-3} - \frac{1}{3n-6} - \frac{n-3}{2n-2} = -\frac{1}{n} - \frac{1}{2(n-2)} + \frac{n-2}{2n-3} - \frac{1}{3(n-2)} - \frac{n-3}{2(n-1)} \\
&= -\frac{1}{n} - \frac{5}{6(n-2)} + \frac{n-2}{2n-3} - \frac{n-3}{2(n-1)} \\
&= \frac{-12(n-1)(n-2)(2n-3) - 10n(2n-3)(n-1)}{12n(n-1)(n-2)(2n-3)} \\
&+ \frac{12n(n-2)(n-1) - (n-3)6n(n-2)(2n-3)}{12n(n-1)(n-2)(2n-3)} = \\
&= \frac{-6(n-1)(n-2)(2n-3) - 5n(n-1)(2n-3)}{6n(n-1)(n-2)(2n-3)} \\
&+ \frac{6n(n-1)(n-2)^2 - 3n(n-3)(n-2)(2n-3)}{6n(n-1)(n-2)(2n-3)} \\
&= \frac{(-6n+6)(2n^2-7n+6) + (-5n^2+5n)(2n-3)}{6n(n-1)(n-2)(2n-3)} \\
&+ \frac{(6n^2-6n)(n^2-4n+4) + (-3n^2+9n)(2n^2-7n+6)}{6n(n-1)(n-2)(2n-3)} = \\
&= \frac{-12n^3 + 42n^2 - 36n + 12n^2 - 42n + 36 - 10n^3 + 15n^2 + 10n^2 - 15n + 6n^4 - 24n^3 + 24n^2}{6n(n-1)(n-2)(2n-3)} \\
&+ \frac{-6n^3 + 24n^2 - 24n - 6n^4 + 21n^3 - 18n^2 + 18n^3 - 63n^2 + 54n}{6n(n-1)(n-2)(2n-3)} \\
&= \frac{-13n^3 + 46n^2 - 63n + 36}{6n(n-1)(n-2)(2n-3)} < 0 \quad \forall n \geq 3
\end{aligned}$$

O valor obtido é um valor negativo pelo que transformar o grafo S_n num outro com o mesmo número de nodos e arestas diminui o seu valor de centralidade de proximidade. Note-se também que basta avaliar situações com um só nodo central porque a existência de um outro nodo central não influencia os termos positivos da soma, apenas aumenta os negativos.

Conclui-se que $\Phi_C(H)$ é máximo se e só se H é S_n .

Calcule-se $\Phi_C(S_n)$.

$$\begin{aligned}
\Phi_C(S_n) &= \sum_{i=1}^n (C_C(x) - C_C(i)) \\
&= (C_C(x) - C_C(1)) + \dots + (C_C(x) - C_C(x)) + \dots + (C_C(x) - C_C(n)) \\
&= \left(\frac{1}{n-1} - \frac{1}{2(n-2)+1} \right) + \dots + \left(\frac{1}{n-1} - \frac{1}{n-1} \right) + \dots + \left(\frac{1}{n-1} - \frac{1}{2(n-2)+1} \right) \\
&= (n-1) \left(\frac{1}{n-1} - \frac{1}{2(n-2)+1} \right) = 1 - \frac{n-1}{2n-3} = \frac{2n-3-n+1}{2n-3} = \frac{n-2}{2n-3}
\end{aligned}$$

4.2.1.3. Centralidade de eficiência

Para maximizar a centralidade de eficiência considere-se de novo $H = S_n$ e averigüe-se se é possível aumentar $\Phi_E(S_n)$ alterando a configuração do grafo.

Acrescentar novas arestas ao grafo não afetará o valor de excentricidade de nenhum dos nodos a menos que um nodo não central passe a ser adjacente a todos os outros, caso em que o centro do grafo passa a ser constituído por dois nodos e uma das parcelas do somatório $\Phi_E(S_n)$ se anula, diminuindo o seu valor.

Ao remover arestas do grafo estrela este ficará desconexo. No entanto, pode averiguar-se o efeito que tem no valor $\Phi_C(S_n)$ retirar uma aresta e introduzir uma outra que mantenha o grafo conexo. Seja x o nodo central, $\{x, y\}$ a aresta removida e $\{y, w\}$ a aresta adicionada ao grafo.

Ora, esta alteração provocará o aumento de uma unidade no valor $ecc(x)$, no valor $ecc(y)$ e no valor $ecc(j) \forall j \in V$, tal que $j \neq x, y, w$. Averigüe-se o efeito destas alterações no somatório $\Phi_C(S_n)$.

Ao valor $\Phi_C(S_n)$ é subtraída $(n - 1)$ vezes a quantidade $\frac{1}{ecc(x)} - \frac{1}{ecc(x)+1}$ e somadas uma vez a quantidade $\frac{1}{ecc(y)} - \frac{1}{ecc(y)+1}$ e $(n - 3)$ vezes a quantidade $\frac{1}{ecc(j)} - \frac{1}{ecc(j)+1}$, sendo j um qualquer nodo de V , tal que $j \neq x, y, w$. Sendo $ecc(x) = 1$, $ecc(y) = 2$, e $ecc(j) = 2$,

$$\begin{aligned} \Phi_C(S_n) - (n - 1) \times \left(\frac{1}{ecc(x)} - \frac{1}{ecc(x) + 1} \right) + \frac{1}{ecc(y)} - \frac{1}{ecc(y) + 1} \\ + (n - 3) \left(\frac{1}{ecc(j)} - \frac{1}{ecc(j) + 1} \right) = \\ \Phi_C(S_n) - (n - 1) \times \left(1 - \frac{1}{2} \right) + \frac{1}{2} - \frac{1}{3} + (n - 3) \times \left(\frac{1}{2} - \frac{1}{3} \right) = \Phi_C(S_n) - \frac{n - 1}{2} + \frac{1}{6} + \frac{n - 3}{6} \\ = \Phi_C(S_n) - \frac{n - 1}{2} + \frac{n - 2}{6} = \Phi_C(S_n) + \frac{-3n + 3 + n - 2}{6} = \Phi_C(S_n) + \frac{1 - 2n}{6}. \end{aligned}$$

Visto que

$$\frac{1 - 2n}{6} < 0 \forall n \geq 1,$$

ao alterar o grafo S_n mantendo o número de arestas, $\Phi_C(S_n)$ diminui. Assim, transformando o grafo S_n num outro com o mesmo número de nodos o seu valor de centralidade de eficiência diminui.

Conclui-se que $\Phi_E(H)$ é máximo se e só se $H = S_n$.

Calcule-se $\Phi_E(S_n)$.

$$\begin{aligned} \Phi_E(S_n) &= \sum_{i=1}^n (C_E(x) - C_E(i)) \\ &= (C_E(x) - C_C(1)) + \dots + (C_E(x) - C_E(x)) + \dots + (C_E(x) - C_E(n)) \\ &= \left(1 - \frac{1}{2} \right) + \dots + (1 - 1) + \dots + \left(1 - \frac{1}{2} \right) = (n - 1) \left(1 - \frac{1}{2} \right) = \frac{n - 1}{2} \end{aligned}$$

4.2.1.4. Centralidade de intermediação

No caso da centralidade de intermediação, a maximização de $\Phi_B(H)$ é bastante simples, visto que o caso ideal é possível, isto é, o caso em que um nodo tem valor máximo de centralidade e os restantes, valor mínimo, ou seja, valor nulo. Tal acontece quando um nodo faz parte de todas as cadeias mais curtas entre quaisquer outros nodos e esses outros nodos não fazem parte de nenhuma cadeia mais curta. Note-se que isto acontece se H for grafo estrela. Resta apenas saber se acontece somente nesse caso.

Ora, se existirem cadeias entre nodos para além, ou em substituição, da cadeia que passa pelo nodo central, a soma $\Phi_B(H)$ diminuirá o seu valor, pois serão acrescentadas quantidades a termos negativos sem que sejam acrescentadas unidades aos termos positivos. Assim, alterações na configuração de $H = S_n$ não melhoram $\Phi_B(H)$. Conclui-se, portanto, que $\Phi_B(H)$ é máximo se e só se $H = S_n$.

Calcule-se $\Phi_B(S_n)$.

$$\begin{aligned}\Phi_B(S_n) &= \sum_{i=1}^n (C_B(x) - C_B(i)) \\ &= (C_B(x) - C_B(1)) + \dots + (C_B(x) - C_B(x)) + \dots + (C_B(x) - C_B(n)) \\ &= \left(\frac{n^2 - 3n + 2}{2} - 0\right) + \dots + \left(\frac{n^2 - 3n + 2}{2} - \frac{n^2 - 3n + 2}{2}\right) + \dots \\ &+ \left(\frac{n^2 - 3n + 2}{2} - 0\right) = (n-1) \left(\frac{n^2 - 3n + 2}{2}\right) = \frac{n^3 - 3n^2 + 2n - n^2 + 3n - 2}{2} \\ &= \frac{n^3 - 4n^2 + 5n - 2}{2}\end{aligned}$$

Resume-se na próxima subsecção os limites superiores de $\Phi_Y(G)$ para cada medida de centralidade Y .

4.3. Sumário

Estudadas as medidas de centralidade de grau, proximidade, eficiência e intermediação em nodos e em grafos, conclui-se que não se verificam relações de implicação entre as várias medidas de centralidade de nodos e, conseqüentemente, entre as características por elas medidas. No que diz respeito às medidas de centralidade de grafos conclui-se que seus limites superiores, em grafos com 2 ou mais nodos, são os apresentados abaixo, sendo estes atingidos se e só se o grafo G é grafo estrela, S_n .

$$\begin{aligned}\Phi_D(G) &\leq n^2 - 3n + 2 \\ \Phi_C(G) &\leq \frac{n-2}{2n-3} \\ \Phi_E(G) &\leq \frac{n-1}{2} \\ \Phi_B(G) &\leq \frac{n^3 - 4n^2 + 5n - 2}{2}\end{aligned}$$

5. Conclusões

5.1. Síntese conclusiva

Na dissertação analisaram-se conjuntos de nodos que são relaxações de cliques e medidas de centralidade em nodos e grafos. Tanto no capítulo 3, Cliques e relaxações de cliques, como no capítulo 4, Medidas de centralidade, os conceitos são apresentados e relacionados por via de resultados todos eles provados na dissertação. É feita em paralelo uma constante interpretação no âmbito das redes sociais. Assim, a dissertação cumpre o seu propósito, resumindo as principais informações dos temas que trata de forma sucinta e simplificada.

É a interpretação no âmbito das redes sociais que une os dois capítulos referidos. Através dos conceitos apresentados no terceiro capítulo são analisados e classificados grupos de pessoas com diferentes características. No quarto capítulo são analisados os vários papéis que um elemento pode ter num grupo e como pode ser medido o quão uniforme é um grupo no que diz respeito a esses papéis. Consegue-se então abranger vários prismas no estudo de um grupo de indivíduos.

5.2. Possíveis desenvolvimentos futuros

As noções expostas no presente trabalho consideram o conceito de distância entre nodos como o número de arestas na cadeia mais curta que os une. A análise desenvolvida na dissertação aplica-se também a grafos em cujas arestas estejam associadas a pesos, grafos em cuja distância entre nodos considere estes pesos. Assim, um possível trabalho futuro seria complementar o presente texto com uma nova secção que contemplasse este tipo de grafos, adaptando as medidas e conjuntos referidos. Repare-se que o contexto das redes sociais deixaria de ser o mais óbvio para interpretar a análise, visto que na representação de uma rede social, se cada aresta simbolizar a existência de uma relação de conhecimento, o seu peso terá sempre valor um. Assim, a interpretação dos resultados teria de ser também adaptada na nova secção. O peso nas arestas dos grafos a analisar poderia simbolizar o número de contactos entre dois indivíduos, por exemplo.

Com base no trabalho desenvolvido na presente dissertação e numa adaptação dos conceitos ao caso em que o peso das arestas dos grafos analisados tome valores superiores a um, poder-se-ão elaborar diversos trabalhos, nomeadamente no contexto marítimo. Estando a signatária de momento a prestar serviço na Marinha, sugerem-se alguns trabalhos neste âmbito.

A Marinha Portuguesa é uma das entidades participantes e competente do sistema integrado de informação e apoio à vigilância, fiscalização e controlo da atividade da pesca¹ (SIFICAP) e por ano as unidades navais realizam aproximadamente 3000 ações de fiscalização a embarcações de pesca a operar em águas sob jurisdição nacional. O sistema SIFICAP em conjunto com o sistema AIS permitem monitorizar a atividade de embarcações de pesca que por lei são obrigadas a operar estes sistemas, o que permite conhecer a sua localização e áreas de atividade. Estas embarcações são, em geral, as que possuem maior dimensão e capacidade de captura, operando em água costeiras e oceânicas.

No âmbito do esforço de fiscalização marítima realizada pela Marinha interessa relacionar as áreas de atividade de pesca com as embarcações que nelas praticam a sua atividade e também com os portos onde o pescado é descarregado, a fim de melhor direcionar o esforço de fiscalização. Neste

¹ Decreto-Lei nº78/2001 de 5 de março.

sentido, interessa conhecer as áreas onde existe maior diversidade e intensidade de atividade de pesca, mas também quais os portos ou lotas praticadas por estas embarcações.

Considerando 3 conjuntos de nodos: A (áreas), E , (embarcações de pesca) e L (lotas) é possível construir um grafo onde uma aresta entre os conjuntos A e E signifique que a embarcação $\{e\}$ pratica a área a e o peso da aresta $\{a, e\}$ corresponda ao nº de dias de pesca realizados num ano. Uma aresta de E para L significa que a embarcação e descarrega o seu pescado na lota l . Neste caso, o peso da aresta $\{e, l\}$ corresponde ao nº de descargas efetuadas num ano. Interessa averiguar quais as áreas com maior diversidade e maior intensidade de pesca e averiguar se existe uma relação entre as lotas e as áreas. Esta análise poderá ser feita com recurso às diferentes medidas de centralidade apresentadas (capítulo 4).

Uma outra aplicação em contexto marítimo do presente estudo é investigação de comportamentos suspeitos no mar. A obrigatoriedade da adoção do sistema AIS através da convenção SOLAS² a navios com mais de 300 ton de deslocamento deu origem a um novo mercado na área das tecnologias de informação e comunicação no âmbito da navegação marítima comercial a nível global. A empresa MarineTraffic regista, por dia, mais de 600 milhões de posições de navios em todo o mundo. Este mercado consiste na comercialização de produtos de análise de grandes volumes de dados provenientes das mensagens AIS enviadas durante as viagens de navios de carga e de passageiros equipados com este sistema. Estes produtos permitem às empresas de transporte marítimo conhecer melhor a sua atividade em termos de eficácia e eficiência assim como conhecer a concorrência. Estes produtos de análise incidem também na deteção de comportamentos suspeitos. Neste caso, o cliente já não são as empresas que fretam navios para transporte de carga ou de passageiros, mas sim, entidades governamentais com responsabilidade na monitorização e controlo da navegação nos espaços marítimos onde têm os seus interesses.

Dois dos principais fornecedores destes serviços de análise são a empresa MarineTraffic com sede na Grécia e a empresa WindWard israelita. No caso da aplicação WindWard são disponibilizadas funcionalidades com recurso à Teoria de Grafos que permitem a análise das relações entre as empresas que são proprietárias de navios, as empresas que operam navios e os próprios navios através de comportamentos suspeitos ocorridos no mar. Desta forma, quando dois navios se encontram no mar em situações onde tal não é suposto acontecer é estabelecido um índice de risco associado a esses navios e o evento é registado para ser posteriormente analisado. Uma aplicação do presente trabalho, em particular, do capítulo 0, é estudo das relações entre empresas através dos comportamentos suspeitos dos navios que estas operam.

² [http://www.mar.ist.utl.pt/mventura/Projecto-Navios-I/IMO-Conventions%20\(copies\)/SOLAS.pdf](http://www.mar.ist.utl.pt/mventura/Projecto-Navios-I/IMO-Conventions%20(copies)/SOLAS.pdf)

6. Bibliografia

- [1]. Moreno, J. (1934). *Who Shall Survive? Foundations of Sociometry, Group Psychotherapy and Sociodrama*. 2, Beacon House. Washington DC.
- [2]. Heider, F. (1946). Attitudes and Cognitive Orientation. *Journal of Psychology*, **21**: 107-112.
- [3]. Cartwright, D. e Harary, F. (1956). Structural Balance: a generalization of Heider's Theory. *Psychological Review*, **63**: 277-293.
- [4]. Harary, F., Norman, R. e Cartwright, D. (1965). *Structural models: An Introduction to the Theory of Directed Graph*. 3, Wiley. Madison.
- [5]. Barnes, J. (1954). Class and committee in a Norwegian island parish. *Human Relations*, **7**: 39-58.
- [6]. Nadel, S. (1957). *The theory of Social Structure*. 1, Routledge. Oxon.
- [7]. Mitchell, J. (1969). *Social Networks in urban Situations: Analyses of Personal Relationships in Central African Towns*. 1, Manchester University Press. Manchester.
- [8]. Lorrain, F. e White, H. (1971). Structural Equivalence of Individuals in Social Networks. *The Journal of Mathematical Sociology*, **1**: 49-80.
- [9]. White, H., Boorman, S. e Breiger, R. (1976). Social structure from multiple networks: I. Blockmodel of roles and positions. *American Journal of Sociology*, **81**: 730-780.
- [10]. Watts, D. e Strogatz, S. (1998). "Collective dynamics of 'small-world' networks". *Nature*, **393**: 440-442.
- [11]. Otte, E. e Rousseau, R. (2002). Social network analysis: a powerful strategy, also for the information sciences. *Journal of information Science*, **28**: 441-453.
- [12]. Luce, R. e Perry, A. (1949). A method of matrix analysis of group structure. *Psychometrika*, **14**: 95-116.
- [13]. Luce, R. (1950). Connectivity and generalized cliques in sociometric group structure. *Psychometrika*, **15**: 169-190.
- [14]. Alba, R. (1973). A graph-theoretic definition of a sociometric clique. *Journal of Mathematical Sociology*, **3**: 113-126.
- [15]. Mokken, R. (1979). Cliques, clubs and clans. *Quality and Quantity*, **13**: 161-173.

- [16]. Swartz, M. (1968). *Local-level Politics: Social and Cultural Perspectives*. 1, Aldine Publishing Company. Chicago.
- [17]. Abello, J., Resende, M. e Sudarsky, S. (2002). Massive Quasi-Clique Detection. *Lecture Notes in Computer Science*, **2286**: 332-343.
- [18]. Seidman, B. e Foster, L. (1978). A graph-theoretic generalization of the clique concept. *Journal of Mathematical Sociology*, **6**: 139-154.
- [19]. Bavelas, A. (1950). Communication Patterns in Task-Oriented Groups. *The Journal of the Acoustical Society of America*, **22**: 725-730.
- [20]. Freeman, C. (1979). Centrality in Social Networks Conceptual Clarification. *Social Networks*, **1**: 215-239.
- [21]. Hage, P. e Harary, F. (1995). Eccentricity and centrality in networks. *Social Networks*, **17**: 57-63.
- [22]. Balasundaram, B. e Pajouh, F. (2013). *Handbook of Combinatorial Optimization*. 1, Springer Science e Business Media. Nova York.
- [23]. Cavique, L., Mendes, A. e Santos, J. (2009). An Algorithm to Discover the k-Clique Cover in Networks. Em: *Portuguese Conference on Artificial Intelligence: Progress in Artificial Intelligence*, Berlim. 363–373.
- [24]. Pattillo, J., Youssef, N. e Butenko, S. (2012). *Handbook of Optimization in Complex Networks*, 1, Springer. Nova York.
- [25]. Quintanilha de Freitas, L. (2010). *Medidas de Centralidade em Grafos*. Tese de Mestrado em Engenharia de Produção. Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia - Universidade Federal do Rio de Janeiro, Rio de Janeiro. 103 pp.
- [26]. Social Science Research Institute (2007). *Closeness Centrality & Betweenness Centrality: A Social Network Lab in R for Beginners*. Acedido em 20 de Novembro de 2019, em: <https://www.youtube.com/watch?v=0aqvVbTyEmc>.
- [27]. Brunato, M., Hoos, H. e Battiti, R. (2007). On Effectively Finding Maximal Quasi-Cliques in Graphs. Em: *International Conference on Learning and Intelligent Optimization: Learning and Intelligent Optimization*, Trento. 41-55.
- [28]. Borba, E. (2013). *Medidas de Centralidade em Grafos e Aplicações em redes de dados*. Tese de Mestrado em Matemática Aplicada. Instituto de Matemática - Universidade Federal do Rio Grande do Sul, Porto Alegre. 66 pp.
- [29]. Systems Inovation (2015). *Network Centrality*. Acedido em 20 de Novembro de 2019, em: <https://www.youtube.com/watch?v=NgUj8DEH5Tc>

- [30]. Ferreira, L. e Gouveia, L. (2020). *Conceitos de Grafos Para Analisar Redes Sociais*. (2020). Texto de apoio à disciplina de Introdução aos Modelos Matemáticos das Licenciaturas em Matemática e Matemática Aplicada. Faculdade de Ciências – Universidade de Lisboa. 44 pp.
- [31]. Harris, J., Hirst, J. e Mossinghoff, M. (2000). *Combinatorics and Graph Theory*, 1, Springer-Verlag. Nova York.
- [32]. Balasundaram, B., Butenko, S. e Hicks, I. (2011). Clique Relaxations in Social Network Analysis: The Maximum k-Plex Problem. *Operations Research*, **59**: 133-142.
- [33]. Shahinpour, S. e Butenko, S. (2013) Algorithms for the maximum k-club problem in graphs. *Journal of Combinatorial Optimization*, **26**: 520–554.