

Análise de Redes Sociais - Página do Público no Facebook

Miguel Sandim e Paula Fortuna

Faculdade de Engenharia da Universidade do Porto,
{miguel.sandim, paula.fortuna}@fe.up.pt

1 Introdução

Com o presente relatório pretende-se apresentar o trabalho desenvolvido sobre Análise de Redes Sociais. Este foi realizado no âmbito da unidade curricular de Extração de Conhecimento de Dados II, recorrendo ao software “Gephi” na sua versão 0.9.1 [1].

O objetivo deste trabalho é analisar uma rede social aplicando várias métricas. Desde logo é importante definir que uma rede social consiste num conjunto finito de atores e nas relações entre eles [2]. Já, por sua vez, a área de Análise de Redes Sociais foca-se nas relações estabelecidas entre os elementos da rede, em vez de na análise dos elementos propriamente ditos.

Mais em particular, neste trabalho foi escolhida para análise a página do Facebook do jornal Público e pretende-se verificar de que forma a Análise de Redes Sociais permite conhecer melhor os posts publicados na página, bem como as comunidades de utilizadores a interagirem com a mesma.

2 Representação da Rede e Avaliação

No presente relatório, optou-se pela representação de uma Rede Social como um grafo não-direcionado $G(V, A)$ constituído por um conjunto de nós (V) e arestas (A). Ao longo do processo de análise da rede foram tidos em conta vários conceitos teóricos (presentes nos *slides* da Unidade Curricular), tendo sido utilizados no trabalho os seguintes:

- Medidas a nível do ator (nó):
 - *Degree* - num grafo não dirigido corresponde ao número de arestas do nó.
 - *Betweenness* - medida que indica quanto é que um nó está entre outros nós da rede.
 - *Closeness* - distância média de um nó a todos os outros.
 - *Eigenvector Centrality* - esta métrica é baseada na ideia de que o poder e estatuto de um ator é definido recursivamente com base no estatuto e poder dos seus vizinhos.
- Medidas ao nível da rede:
 - *Geodesic distance* - comprimento do caminho mais curto entre dois nós.
 - *Eccentricity* - maior *geodesic distance* entre um nó e qualquer outro da rede.
 - *Diâmetro* - o maior de todos os caminhos mais curtos entre todos os pares de nós na rede.
 - *Raio - eccentricity* mínima de todos os nós da rede.
- Outros conceitos:
 - *Bridge* - aresta que conecta dois nós, sem a qual não é possível encontrar um caminho que una os dois.
 - *Gatekeepers* - nós que se ligam a uma *bridge*.
 - *Community Detection* - Detecção de nós fortemente conectados entre si, e pouco conectados com os nós das restantes comunidades.

3 Recolha de dados

O conjunto de dados utilizado corresponde ao grafo da interação entre Utilizadores e Publicações da página do Público, recolhidos através da ferramenta “Netvizz”. A recolha foi realizada no dia 9 de Maio de 2016 e inclui dados relativos às 50 últimas publicações da página, bem como de todos os comentários a estas. Esta plataforma fornece o grafo num ficheiro no formato de grafo “.dgf” que a ferramenta “Gephi” permite importar.

Assim, no contexto desta rede, os nós correspondem a Utilizadores e a Publicações e as arestas a interações entre um Utilizador e uma Publicação (incluindo *likes* e comentários a esta). No caso das arestas, o peso destas corresponde ao número de interações entre o Utilizador e a Publicação em questão.

Com uma breve análise em R concluiu-se que o estudo foi realizado com nós referentes a 12393 Utilizadores diferentes e a 50 Publicações (48 partilhas de *links* e 2 partilhas de fotografias). A rede possui também 16436 arestas, com uma média de 1.33 interações por Utilizador (existem 2162 utilizadores com mais do que 1 interação) e um peso médio de 1.03 (existem 300 utilizadores com uma média de pesos das suas interações maior do que 1).

4 Análise da Rede

A análise da rede consistiu numa série de passos realizados no “Gephi”, tais como Identificação de *Layout*, Extração de Comunidades e Extração de Métricas da Rede, que se apresentam nos próximos subcapítulos.

4.1 Identificação de *Layout*

Para identificação do *Layout* mais adequado à rede em estudo, foram utilizados vários algoritmos. Dada o número elevado de nós e arestas não foi encontrado de forma imediata um *layout* que permitisse ler de forma clara a rede. O melhor resultado foi obtido recorrendo ao algoritmo “Atlas Force 2”.

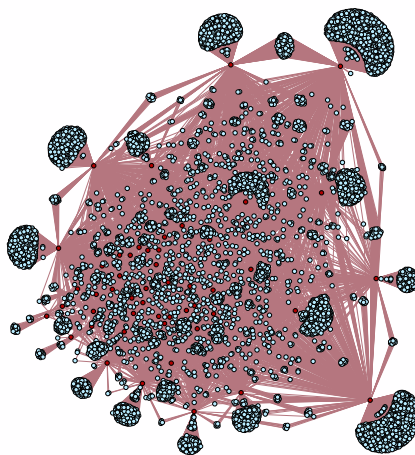


Figura 1. Layout obtido utilizando-se o algoritmo “Atlas Force 2” – a azul encontram-se os nós de utilizadores e a vermelho os de publicações.

Tal como pode ser verificado na figura 1, dado o elevado número de nós e arestas torna-se difícil identificar grupos nesta rede. Contudo, ao mesmo tempo verifica-se um padrão: para cada nó do tipo “publicação” (vermelho), existe um elevado número de utilizadores que se liga apenas a esse nó. Dado que esse número de utilizadores é tão elevado, a rede reflete sobretudo esse efeito, o que é reforçado na figura 2 ao se analisarem as comunidades extraídas (o processo de extração de comunidades será analisado numa secção posterior).

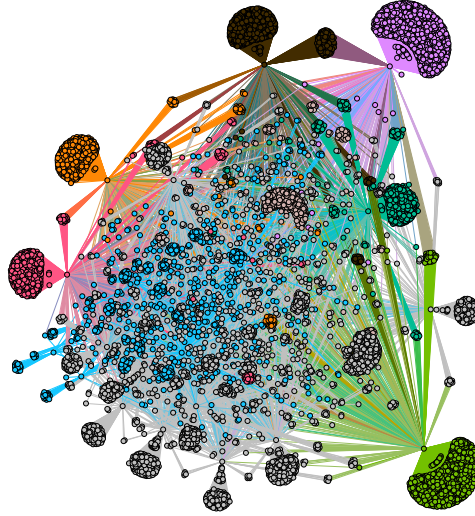


Figura 2. *Layout* obtido utilizando-se o algoritmo “Atlas Force 2” e *Modularity* para extração de comunidades. Cada cor representa uma comunidade, enquanto o cinzento representa as várias comunidades de menores dimensões.

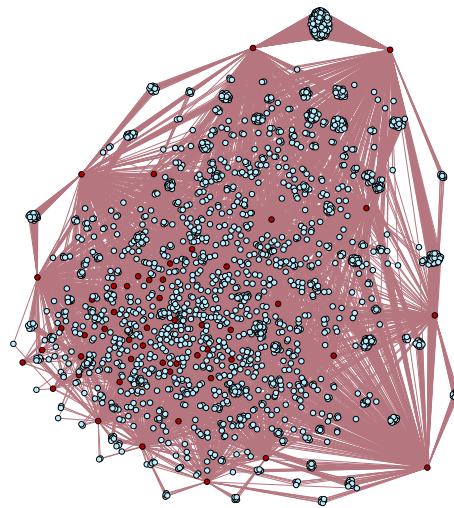


Figura 3. Rede simplificada, sem nós com *degree* igual a um.

Tal como pode ser visto na figura 2, as comunidades obtidas resumem-se em grande parte a utilizadores que responderam a uma e só uma publicação. Neste caso verifica-se um grande número de arestas que funcionam como *bridges* e cada publicação como *gatekeeper*.

Contudo, seria também interessante explorar se, para além deste efeito principal, existem padrões na rede para os utilizadores que responderam a várias publicações. Neste sentido, foram eliminados da rede os nós com *degree* de um. Para isso foi aplicado um filtro *Degree Range* em que o limite inferior é dois. O resultado desta operação é apresentado na figura 3.

Com o novo número de nós encontrado, uma vez que a rede continua bastante densa, o algoritmo “Noverlap” permitiu encontrar o *layout* mais perceptível. Apresenta-se na figura 4 o *layout*, com análise do *degree* de cada um dos nós.

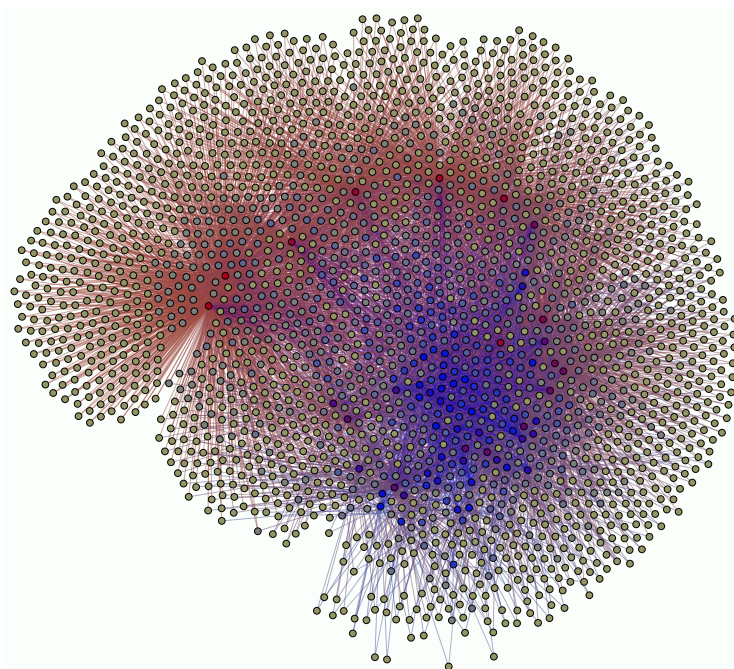


Figura 4. Rede em que o gradiente de cor dos nós corresponde a diferentes *degrees*. Ao vermelho corresponde um maior *degree*, ao azul intermédio e ao amarelo menor.

Como é possível verificar na figura 4 a maior parte dos utilizadores tem um *degree* baixo (amarelo). Por sua vez, a maior parte das publicações apresenta um *degree* elevado (vermelho). Existe ainda uma terceira zona da rede onde tanto utilizadores como publicações têm um *degree* intermédio (azul).

4.2 Extração de Comunidades

Para além do que foi apresentado até aqui, procurou-se também extrair comunidades, recorrendo a uma medida “Modularity” disponibilizada no “Gephi” (Método de *Louvain*). Esta indica quão bem uma rede se decompõe em comunidades. Um valor mais elevado nesta métrica indica uma estrutura interna mais sofisticada. O resultado obtido apresenta-se na figura 5.

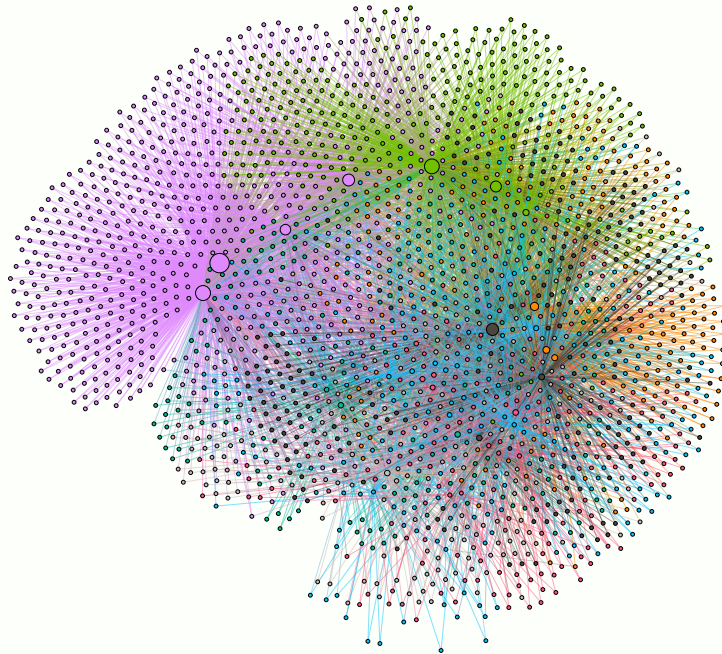


Figura 5. Rede em que cada cor dos nós corresponde a uma comunidade e em que o tamanho dos nós é proporcional à medida de *betweenness*.

Depois disso para melhor se compreenderem as comunidades presentes procedeu-se a nova organização de *layout* e analisou-se o conteúdo das publicações com maior *betweenness*, tendo-se obtido a rede presente na imagem 6.

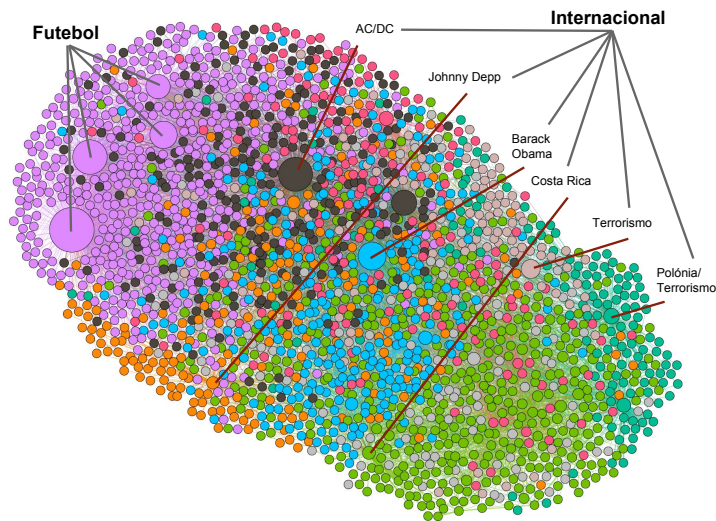


Figura 6. Comunidades presentes na rede, bem como o tema da publicação com maior *betweenness* nessa comunidade.

Na figura 6 é apresentada o *layout* da rede quando se utiliza o algoritmo “Force Atlas 2”. Foi escolhido este algoritmo porque é o que facilita a compreensão das várias comunidades encontradas.

Relativamente aos resultados obtidos, pode concluir-se que existe uma comunidade que se distingue das restantes – esta está representada a roxo na figura 6 . É de notar que todas as publicações nesta comunidade se referem a futebol (e.g. Benfica, Cristiano Ronaldo). Por sua vez, nas restantes comunidades nenhum dos algoritmos de *layouts* permitiu separar claramente as comunidades entre si. Assim, analisou-se o assunto da publicação com maior *betweenness* em cada comunidade e gerou-se, para cada um deles, uma *label* (apresentadas na figura 6). Podemos concluir que as restantes comunidades podem ser aglomeradas em Internacional (e.g. Costa Rica, Barack Obama).

4.3 Outras Métricas

Tabela 1. Estatísticas sobre a rede.

Statistic Group	Value
Degree	
- Average Degree	5.61
Betweenness	
- Diameter	5
- Radius	3
- Avg Path Length	3.28
Modularity	
- Modularity	0.38
- Number of communities	10
Eigenvector Centrality	
- Sum Change	0.02

Ao longo do procedimento seguido no “Gephi” foram sendo obtidas diferentes métricas em diversos relatórios. Estas estão sumariadas na tabela 1.

5 Conclusão

Com o presente trabalho pretendia-se analisar uma rede social e para isso foram recolhidos dados sobre as publicações na página do Facebook do jornal Público.

Dado o número elevado de publicações e comentários às mesmas, a rede obtida apresentava uma grande densidade de nós e arestas, dificultando a sua análise visual, que é o ponto forte do Gephi. Desde logo, numa primeira análise salientava-se que cada publicação tinha uma comunidade de utilizadores que só respondera a essa publicação. Para tentar aprofundar outros efeitos que pudessem estar a ocorrer procurou-se simplificar a rede, retirando os nós ligados apenas a uma publicação. Depois de se proceder a este processamento foi possível concluir que existe uma comunidade de utilizadores ligada a publicações sobre futebol e também outra comunidade de utilizadores ligada a publicações de temas mais variados e internacionais. Contudo, ao mesmo tempo é de salientar que estas comunidades não são mutuamente exclusivas.

Para além disso, este trabalho permitiu ainda concluir que, através de extração de comunidades, é possível chegar a um resultado semelhante a técnicas de *clustering*. Contudo, salienta-se também que esta técnica não necessita de conhecer as características dos nós para o fazer, o que hoje em dia é vantajoso na análise de redes sociais, dado que as APIs de várias redes sociais (e.g. Facebook, Twitter) têm vindo a restringir cada vez mais as informações disponibilizadas. Por exemplo, no caso dos utilizadores, o Facebook fornece apenas o ID e o seu nome.

Em jeito de conclusão, foram cumpridos os objetivos do presente trabalho, uma vez que foi analisada visualmente uma rede social. Através desta análise foi possível chegar a várias informações e conclusões sobre a rede.

Referências

1. Mathieu Bastian, Sebastien Heymann e Mathieu Jacomy. *Gephi: An Open Source Software for Exploring and Manipulating Networks*. 2009. URL: <http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154>.
2. Stanley Wasserman e Katherine Faust. *Social network analysis: Methods and applications*. Vol. 8. Cambridge university press, 1994.