

# DETECTABILIDADE DE GRUPOS ESTRATÉGICOS EM REDES SOCIAIS

Ana Carolina Wagner G. de Barros

Orientador: Moacyr Alvim

acwgdb@gmail.com



24 de maio de 2018

## 1 INTRODUÇÃO

- Contexto
- Problema
- Justificativas e Objetivo

## 2 REFERENCIAL TEÓRICO

## 3 METODOLOGIA

## 4 RESULTADOS

## 5 PRÓXIMOS PASSOS

## 6 REFERÊNCIAS

# INTRODUÇÃO

## Contexto

- Prestígio e sucesso científico de um pesquisador:
  - quantidade de publicações ("*publish or perish*");
  - o número de citações recebidas: indicação da relevância e qualidade de um artigo;
  - relevância dos periódicos nos quais os artigos produzidos são publicados.
- Dinâmica de produção acadêmica modelada como uma rede social.
  - A dinâmica de crescimento é regida pelo princípio da **anexação preferencial** (BIANCONI; BARABÁSI, 2001);
  - Grande desafio por parte dos **nós periféricos** em conseguir ganhos de centralidade (medida em grau).
- **Comportamento estratégico:** buscando conquistar um maior prestígio acadêmico (ganhos de centralidade), os periódicos classificados como **periféricos estratégicos** fogem da lógica da ligação preferencial e iniciam o processo de **reciprocidade periférica**, em que fazem citações somente entre os periódicos pertencentes ao próprio grupo crítico que estão inseridos (SANDE, 2016).

# INTRODUÇÃO

## Problema

- A manipulação de uma lógica de anexação de nós periféricos com baixa centralidade pertencentes a um grupo específico (grupo estratégico), evitando a lógica de ligação preferencial, favorece outros nós pertencentes ao mesmo grupo à custa de nós com maior centralidade.
- Os nós que adotam o **comportamento estratégico**:
  - continuam a ser vistos pelo restante da rede como nós comuns;
  - podem ser vistos como um **grupo estratégico** e os outros nós da rede como um **grupo não estratégico**;
  - não podemos afirmar que tal grupo estratégico forme necessariamente uma comunidade; em vez disso, eles estão inseridos na comunidade como **nós periféricos**.
- Procurar uma modificação dos métodos de identificação de comunidades para que seja possível identificar o surgimento de grupos estratégicos.

**Podemos detectar o comportamento estratégico?**  
**Como podemos identificar o surgimento desses grupos estratégicos?**

# INTRODUÇÃO

## Justificativas e Objetivo

- **Rede socialmente eficiente:** maximiza a soma total da utilidade, ou seja, maximiza o bem-estar geral (JACKSON, 2008).
- No entanto, o **comportamento estratégico** diminui essa eficiência:
  - prioriza o aumento de um ganho local em detrimento de um ganho global;
  - Redes de citações: as citações devem refletir o potencial acadêmico de fato.
- Portanto, conseguindo detectar a ocorrência do comportamento estratégico em tais redes podemos tentar evitar fraudes.

**Objetivo:** Propor um algoritmo de detecção de grupos estratégicos e uma medida que avalie a eficiência da rede.

Noções teóricas estudadas para a compreensão e realização deste trabalho:

- Uma **comunidade** é um subgrafo que contém nós que estão mais densamente ligados uns aos outros do que ao resto da rede ou, equivalentemente, um grafo tem uma estrutura de comunidade se o número de links em qualquer subgrafo for maior que o número de links entre esses subgrafos (NEWMAN; GIRVAN, 2004).
- Um método clássico para detectar comunidades é encontrar uma partição do conjunto de nós que maximiza uma função de otimização chamada **modularidade** (NEWMAN, 2003). Para uma dada divisão dos vértices da rede em alguns módulos, a modularidade reflete a concentração de arestas dentro de módulos comparada com a distribuição aleatória de arestas entre todos os nós, independentemente dos módulos.(DUGUÉ; PEREZ, 2015).
- A evolução da rede também é impactada pelos atributos dos nós (KIM; ALTMANN, 2015).

- A probabilidade de uma citação feita por um artigo para outro é de fato uma função crescente da similaridade entre os dois artigos. A criação de links nas redes de citações pode ser vista como governada pela **homofilia** (CIOTTI et al., 2016).
- A **homofilia** promove um grupo de nós com propriedades comuns a serem mais densamente integrados do que os grupos sem essa propriedade.
- O comportamento homofílico combinado com diferenças de tamanho de grupo pode colocar os grupos minoritários em desvantagem, limitando sua visibilidade (medida por grau) nas redes sociais (KARIMI et al., 2017).
- Indícios de **comportamento assimétrico e homofílico** nas redes de citações (KARIMI et al., 2017).

- Inicialmente, por meio do *Stochastic Block Model* (SBM), é gerado um grafo aleatório com dois grupos (1-estratégico e 2-não estratégico) e calculado dois valores de submodularidade propostos ( $Q_1$  e  $Q_2$ ) para confirmar que **o comportamento estratégico não pode ser identificado pelo conceito de modularidade** (BARROS et al., 2017).

$$Q_1 = \frac{1}{m} s_1^T B s_1 \quad Q_2 = \frac{1}{m} s_1^T B s_2$$

- Em seguida, como proposta é sugerido que para identificar esse comportamento específico, é necessário observar a rede em diferentes intervalos de tempo para verificar a variação do número de links dentro e entre os grupos, considerando que o número total de links em o gráfico permanece constante. Para isso, são utilizadas as seguintes razões:

$$\alpha = \frac{c_{21}(t)}{c_{11}(t)} \quad \alpha^+ = \frac{c_{21}(t + \Delta t)}{c_{11}(t + \Delta t)} \quad \beta = \frac{c_{12}(t)}{c_{22}(t)} \quad \beta^+ = \frac{c_{12}(t + \Delta t)}{c_{22}(t + \Delta t)}$$

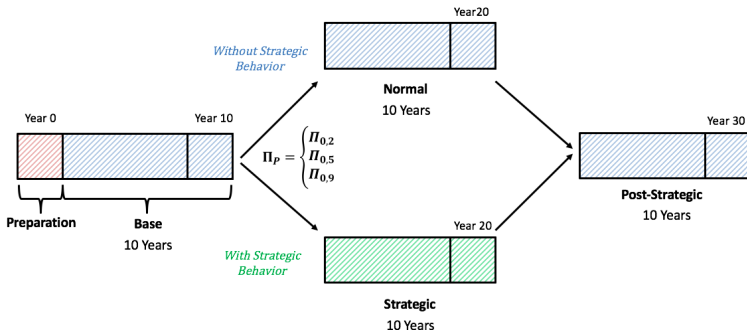
$$R_1 = \frac{\alpha}{\alpha^+} \quad R_2 = \frac{\beta}{\beta^+}$$

onde  $c_{ij}$  representa o número de links do grupo  $j$  para  $i$  ( $i, j = 1, 2$ ).



- **Comunidade:**  $R_1 > 1$  e  $R_2 > 1$ 
  - Se tivermos duas comunidades normais, espera-se que o número de links em ambos os grupos aumente e entre eles diminua.
- **Grupo estratégico:**  $R_1 > 1$  e  $R_2 \approx 1$ 
  - Se tivermos um grupo estratégico, os links dentro desse grupo aumentarão e os links do grupo 1 para o grupo 2 diminuirão na mesma proporção. Além disso, os links do grupo 2 para o grupo 1 e dentro do grupo 2 tenderão a permanecer os mesmos, porque o grupo 2 continuará a ver o resto da rede como nós aleatórios.
- A validação dos resultados é realizada, inicialmente, a partir de uma rede de citação simulada (SANDE, 2016), em seguida, usando uma rede de citações real gerada a partir dos dados extraídos da base do *Web of Science*.

- Rede de citações simulada (400 periódicos):

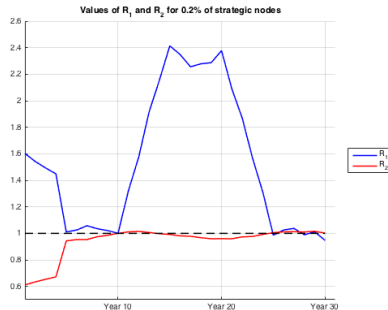
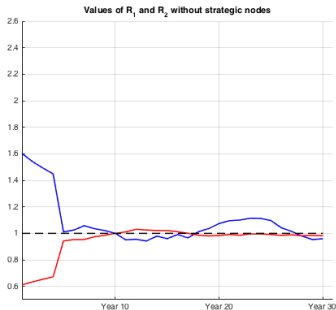


- Temos duas redes comparáveis: (1) a que segue a lógica da anexação preferencial e (2) a que adota o comportamento estratégico (foi escolhido  $p = 20\%$ ).
- Para cada uma dessas redes, uma matriz  $A$  é extraída ao final do ano 10 e para cada um dos 30 anos de simulação são calculadas as razões  $R_1$  e  $R_2$ .

- Rede de citações gerada a partir do WoS:
  - Área escolhida: *Mathematical & Computational Biology*
  - Período de Publicação: 2010 a 2017
  - **(1)** Para cada um dos 59 periódicos ativos durante esses anos foram extraídos todos\* os artigos publicados nesse período; **(2)** Para cada um desses artigos ( $\approx 50$  mil) foram extraídos os artigos citados, totalizando mais de 1,5 milhões de citações.
  - <[https://github.com/anacwagner/WOS\\_CitationNetworks](https://github.com/anacwagner/WOS_CitationNetworks)>

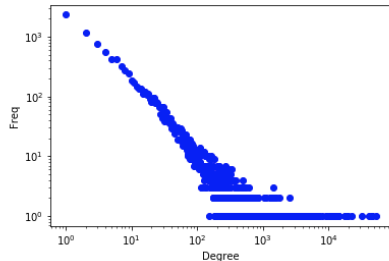
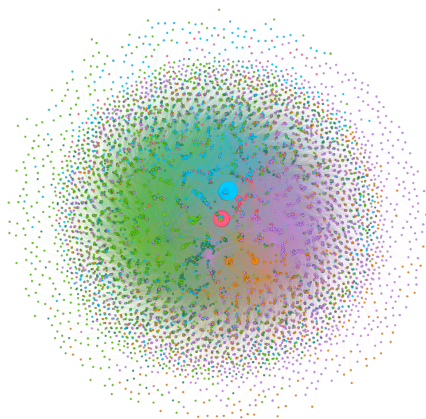
# RESULTADOS

- Razões  $R_1$  e  $R_2$  calculadas para uma rede de citações simulada.
- Mostrou-se eficaz na identificação de comunidades e também no comportamento estratégico, se esse comportamento é previamente conhecido.



# RESULTADOS


- Rede de citações real extraída a partir do WoS.
  - Grau médio: 9,05
  - A análise de modularidade identificou 6 grupos.




# PRÓXIMOS PASSOS

- Usar a rede real extraída do WoS para testar as razões propostas  $R_1$  e  $R_2$  (validar a medida e a rede simulada).
- Aplicar o código da simulação feita sobre a rede real. A rede real vai funcionar como a base e continuamos a simulação a partir disso (alguns parâmetros precisam ser adaptados).
- A partir da análise de modularidade do Gephi sobre a rede real, podemos identificar 6 grupos. Um deles será o grupo com os nós periféricos e o restante o grupo normal.

# REFERÊNCIAS I

 BARROS, A. C. W. G. et al. Analyzing the formation of groups in a network adapting the modularity concept. In: *I Simpósio Internacional Network Science*. [S.l.: s.n.], 2017.

 BIANCONI, G.; BARABÁSI, A.-L. Competition and multiscaling in evolving networks. *Europhysics Letters*, v. 54, n. 4, p. 436–442, 2001.


 CIOTTI, V. et al. Homophily and missing links in citation networks. *EPJ Data Science*, v. 5, n. 1, p. 7, Mar 2016. ISSN 2193-1127. Disponível em: <<https://doi.org/10.1140/epjds/s13688-016-0068-2>>.


 DUGUÉ, N.; PEREZ, A. *Directed Louvain : maximizing modularity in directed networks*. [S.l.], 2015.


 JACKSON, M. O. *Social and Economic Networks*. Princeton, NJ, USA: Princeton University Press, 2008.


 KARIMI, F. et al. Visibility of minorities in social networks. *CoRR*, abs/1702.00150, 2017.

# REFERÊNCIAS II

 KIM, K.; ALTMANN, J. *Effect of Homophily on Network Evolution*. [S.l.], 2015. Disponível em: <<https://ideas.repec.org/p/snv/dp2009/2015121.html>>.

 NEWMAN, M. E. J. The Structure and Function of Complex Networks. v. 45, n. 2, p. 167–256, 2003.

 NEWMAN, M. E. J.; GIRVAN, M. Finding and evaluating community structure in networks. p. 1–16, 2004.

 SANDE, W. W. C. *Reciprocidade periférica como estratégia para aumento de centralidade: estudo de rede de citações acadêmicas*. Tese (Doutorado) — Ebape-FGV, 2016.