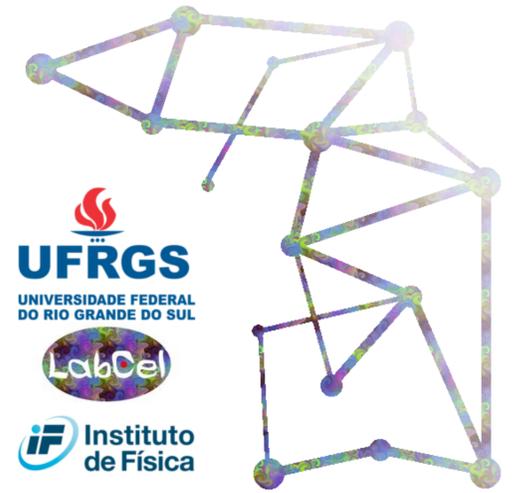


Equação-mestra para a distribuição de grau de uma rede de Duplicação e Divergência

Vítor Sudbrack (IC), Daniel Gamermann (PhD)

Instituto de Física, UFRGS

vitor.sudbrack@ufrgs.br



Resumo

Crescimento de redes descritas pelo modelo de Duplicação e Divergência propõe um ideia genérica simples de dinâmica natural de redes. Em particular, é uma alternativa para o modelo conhecido de Barabási-Albert quando aplicadas às redes de proteína.

Nesse trabalho nós derivamos a equação mestra para a distribuição de grau dos nós de uma rede que cresce conforme dinâmica de Duplicação e Divergência e foi obtida uma expressão para o número total de ligações como uma função do número de nodos. Usando ferramentas da álgebra linear foi investigado o comportamento assintótico da distribuição de grau.

Os resultados analíticos mostram que o grau médio das redes convergem se a taxa de mutação total é maior que 0.5 e diverge caso contrário.

Comparamos as soluções obtidas com simulações numéricas. Esses resultados são compatíveis com os valores esperados a partir das equações deduzidas. O método desenvolvido é uma ferramenta robusta para investigar outras classes de dinâmica de redes.

Introdução

Como a física engloba o estudo e descrição de fenômenos complexos em áreas distintas, como biologia ou ciências sociais, a **teoria de grafos** é comumente aplicada. Nesta abordagem o sistema estudado é descrito como uma rede, que consiste de um conjunto de **nós** e um conjunto de **ligações** entre eles.

Neste trabalho, estamos interessados em **dinâmicas de redes**, isto é, o estudo de uma dinâmica por trás de um grafo que se expande. A motivação para tal estudo está na dinâmica de redes biológicas. Muitos sistemas biológicos encontram uma descrição natural dentro desse modelo e entendê-lo com as regras da seleção natural que resulta em redes com estruturas topológicas parecidas com as redes reais observadas nos fornece ideias sobre a dinâmica da evolução das espécies.



Figura 1: Aqui, usamos uma descrição de processo markoviano para o modelo de dinâmica de redes. Nessa abordagem, o estado de uma rede num dado instante depende apenas na sua configuração no momento anterior. Dado um conjunto de regras que descrevem como a rede muda em cada passo de tempo, construímos a correspondente equação mestra que representa a evolução das configurações do sistema.

A importância de descrever o comportamento médio de um processo estocástico dessa maneira está em poder saber o comportamento da rede para diferentes valores de parâmetros sem a necessidade de longos tempos de simulação numérica para se obter informações estatisticamente relevantes.

Modelo de dinâmica: Duplicação e divergência

Dado uma pequena rede inicial (três nós conectados uns aos outros formando um triângulo), estudamos o processo markoviano em que, em cada instante de tempo, repetimos o seguinte processo:



Figura 2: Escolhemos um nó ao acaso. Adicionamos na rede um idêntico (com as mesmas ligações). Cada ligação do nó original é perdida com uma probabilidade m_o e cada ligação do nó duplicado é perdida com uma probabilidade m_d . Também, uma ligação entre os nós original e duplicado é sempre criada.

Este modelo é uma adaptação do modelo proposto por Vázquez *et al.*. Vázquez baseou seu no crescimento de redes com **regras locais**, isto é, regras que requerem informação somente de um nó ao invés de outras regras que demandam informação sobre toda a rede. Esse modelo de dinâmica de redes permite que todas as proteínas possam evoluir de um ancestral comum através de **cópia genética** (representada pela duplicação) e **mutações** (divergência). Portanto, esse modelo se assemelha toda a história de um evolução do genoma.

Resultados

Grau médio e número total de ligações

Primeiro, analisamos o comportamento do grau médio da rede, \bar{k} . Suponha que um nó de grau médio tenha sido escolhido para ser duplicado. No instante seguinte, o número de ligações é dado por:

$$L_{t+1} = L_t + \bar{k} - m_o \bar{k} - m_d \bar{k} + 1. \quad (1)$$

Os termos no lado direito da Eq. (1) representam, da esquerda para direita: as ligações já existentes, os nós adicionados devido a duplicação de um nó de grau médio, o número médio de ligações perdidas do nó original, o número médio de ligações perdidas do nó cópia e a criação da ligação cópia-copiado.

A soma das mutações (parâmetros de divergência) é definida como a **mutação total**, $M = m_o + m_d$. Agora, dado o *handshake lemma*, $2L = \sum_{i=1}^N k_i = N\bar{k}$, podemos encontrar uma relação entre o grau médio em dois instante da rede. Para longos tempos, podemos aproximar o mapa para a seguinte equação diferencial contínua,

$$\frac{\partial \bar{k}}{\partial t} = \frac{1 - 2M}{t + 1} \bar{k} + \frac{2}{t + 1} \quad (2)$$

Para $0 \leq M \leq 2$, a solução da EDO acima é dada por

$$\bar{k}(t) = \begin{cases} c_0(t+1)^{1-2M} + \frac{2}{2M-1} & \text{para } M \neq 0.5; \\ 2 \log(t+1) + c_0 & \text{para } M = 0.5, \end{cases} \quad (3)$$

$$L(t) = \begin{cases} c_1 t(t+1)^{(1-2M)} + \frac{t}{2M-1} & \text{para } M \neq 0.5; \\ t \log(t+1) + c_1 t & \text{para } M = 0.5 \end{cases} \quad (4)$$

onde as constantes c_i estão relacionada com as condições iniciais da rede.

Interessante notar o comportamento assintótico $t \rightarrow \infty$ do grau médio dado por Eq. (3). Para $M > 0.5$, o grau médio converge para $\frac{2}{2M-1}$. Para $M < 0.5$, o grau médio diverge como t^{1-2M} e logaritmicamente para $M = 0.5$.

Simulações computacionais mostram que redes com 20 mil nós já atingiram o grau médio estacionário para aprox. $M > 0.8$, compatíveis com o esperado pela Eq. (3). Quanto mais nos aproximamos do ponto de transição de comportamento, maior devem ser as redes para observarmos o grau médio estacionário.

Distribuição de grau em tempos finitos

Considerando todas as possíveis ocorrências e suas probabilidades, é possível escrever a relação entre o número de nós com grau k , $N(k, t)$, no instante t e $t + 1$:

$$\begin{aligned} N(k, t+1) = & N(k, t) - \frac{N(k, t)}{t} \\ & + \sum_{i=0}^{t-1} P_o(i \rightarrow k) \frac{N(i, t)}{t} + \sum_{i=0}^{t-1} P_d(i \rightarrow k) \frac{N(i, t)}{t} \\ & + \frac{(k-1)N(k-1, t)(1-m_o)(1-m_d)}{t} \\ & + \frac{(k+1)N(k+1, t)m_o m_d}{t} \\ & - \frac{kN(k, t)[(1-m_o)(1-m_d) + m_o m_d]}{t} \end{aligned} \quad (5)$$

os termos nessa equação representam, em ordem: o número existente de nós de grau k ; a possibilidade de um nó de grau k ser escolhido para ser duplicado; o primeiro somatório reflete a probabilidade de um nó original de grau i sofrer duplicação e se tornar de grau k , o termo análogo para o grau duplicado; por fim, os últimos três termos referem-se a probabilidade de que um nó vizinho seja escolhido para ser duplicado, vindo dos graus $k-1$ ou $k+1$ para k , e o fluxo oposto. As probabilidades dentro dos somatórios são dadas por distribuições binomiais que representam a probabilidade, tanto para o nó original quanto cópia, de ir de um grau i para k :

$$P_o(i \rightarrow k) = C_{k-1}^i (1-m_o)^{k-1} (m_o)^{i-k+1} \quad (6a)$$

$$P_d(i \rightarrow k) = C_{k-1}^i (1-m_d)^{k-1} (m_d)^{i-k+1} \quad (6b)$$

Como mostra a Fig. 3, a solução numérica do mapa da Eq. (5) produz excelentes resultados quando comparado com médias sobre simulações numéricas do processo estocástico.

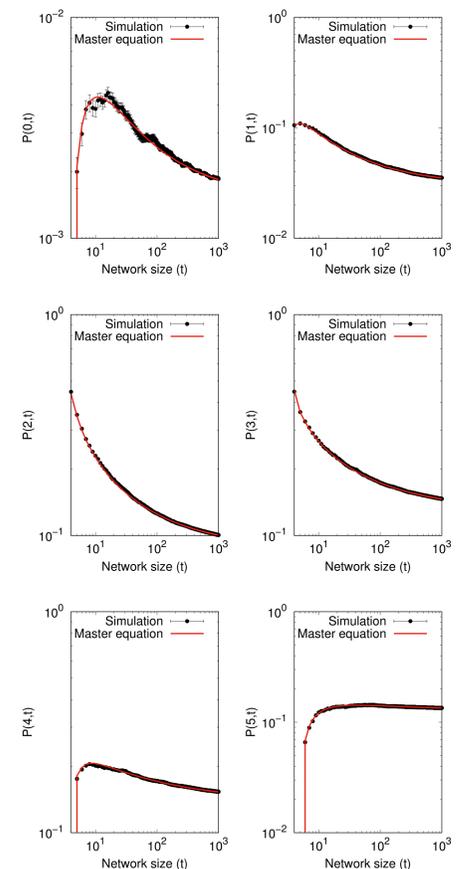


Figura 3: Comparação dos resultados da fração média de nós com grau k em simulações de 5 mil redes começando com 3 nós até 10^3 e resultados analíticos da integração numérica da equação 5 para $m_o = 0.25$ e $m_d = 0.40$, e, portanto, $M = 0.65$.

Distribuição de grau no limite $t \rightarrow \infty$

A Eq. (5) pode ser convenientemente escrita na forma matricial, definindo um vetor coluna $\vec{N}(t)$, onde $N_k(t) = N(k, t)$. É possível escrever $\frac{d}{dt} \vec{N} = \frac{1}{t} A \vec{N}$, onde a matriz A apresenta componentes constantes no tempo.

Normalizando o vetor $\vec{N}(t)$, para se obter uma quantidade que não divirja no limite estudado, definimos $\vec{P}(t) = \frac{1}{t} \vec{N}(t)$, que pode ser interpretada como a fração de nós de grau k no instante t . A solução da equação matricial, pode então ser escrita como a combinação linear dos autovetores da matriz A , $\vec{P} = \sum_{i=1}^t k_i \vec{X}_i t^{\lambda_i - 1}$.

Quando existe um estado estacionário, então haverá um autovalor $\lambda = 1$ para a matriz A e este será o maior autovalor (caso contrário, P iria divergir). Este autovetor associado ao autovalor 1 é dito autovetor principal e será a distribuição de grau estacionária da rede.

Conclusão

Nós derivamos a equação mestra para a rede que cresce via duplicação e divergência. A integração numérica da equação forneceu ótimos resultados quando comparada com a média das simulações numéricas para todos os valores de mutação total.

Nós buscamos uma solução numérica assintótica, que também concorda com os dados simulados para valores de mutação maiores que 0,5, quando é esperado que a distribuição de grau converja para uma solução assintótica.

Referências

- [1] U. Bhat, P. L. Krapivsky, R. Lambiotte, and S. Redner. Den-sification and structural transitions in networks that grow by node copying. *Phys. Rev. E*, 94:062302, Dec 2016.
- [2] Alexei Vázquez. Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Phys. Rev. E*, 67:056104, May 2003.
- [3] Alexei Vázquez, Alessandro Flammini, Amos Maritan, and Alessandro Vespignani. Modeling of protein interaction networks. *Complexus*, 1(1):38–44, 2003.