



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
DEPARTAMENTO DE FÍSICA
CURSO DE GRADUAÇÃO EM FÍSICA

MANOEL LOURENÇO

**GERANDO VARIÁVEIS ALEATÓRIAS DISCRETAS COM DISTRIBUIÇÃO EM LEI
DE POTÊNCIA**

FORTALEZA

2019

MANOEL LOURENÇO

GERANDO VARIÁVEIS ALEATÓRIAS DISCRETAS COM DISTRIBUIÇÃO EM LEI DE
POTÊNCIA

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Física do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Física.

Orientador: Prof. Dr. André Auto Moreira

Coorientador: Dr. Rilder de Sousa Pires

FORTALEZA

2019

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

A48g Alves Neto, Manoel Lourenço.
Gerando variáveis aleatórias discretas com distribuição em de Lei de Potência : Método da inversão /
Manoel Lourenço Alves Neto. – 2019.
34 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Centro de Ciências,
Curso de Física, Fortaleza, 2019.

Orientação: Prof. Dr. André Auto Moreira.

Coorientação: Prof. Dr. Rilder de Sousa Pires.

1. Lei de potência. 2. grafos. 3. network science. I. Título.

CDD 530

MANOEL LOURENÇO

GERANDO VARIÁVEIS ALEATÓRIAS DISCRETAS COM DISTRIBUIÇÃO EM LEI DE
POTÊNCIA

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Física do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Física.

Aprovada em:

BANCA EXAMINADORA

Prof. Dr. André Auto Moreira (Orientador)
Universidade Federal do Ceará (UFC)

Dr. Rilder de Sousa Pires (Coorientador)
Instituto de Pesquisa e Estratégia Econômica do Ceará
(IPECE)

Prof. Dr. Marcos Antônio Araújo Silva
Universidade Federal do Ceará (UFC)

À minha minha mãe e irmã por serem as mulheres mais fortes que conheço e meus filhos Davi e Pedro.

AGRADECIMENTOS

Ao grande mestre e amigo André Auto por me orientar e conduzir do caminho da obscuridade à luz da sapiência.

Ao grande amigo Rilder Pires por me ensinar o que é gerador de números aleatórios e me introduzir a complexidade.

Aos amigos Marciel, Israel, Johnathan Sales, Wagner Sena do grupo Sistemas Complexos por sempre me incentivarem e ajudarem a esclarecer minhas dúvidas.

Aos grandes amigos Ramon Sampaio, Daniel de Araújo e Lucas Sievers (*Matuza*) por me apoiarem e proporcionarem discussões sobre física, ciência, filosofia, sociologia, antropologia, política, cerveja e culinária.

À minha mãe e irmã por sempre me apoiarem e acreditarem na capacidade de superar dificuldades e desafios cujo o apoio e amor foi de fundamental importância para concluir esse curso.

À todos os funcionários e colaboradores do departamento de física por proporcionarem um ambiente limpo e funcionando para todos que usufruem desse departamento.

À todos os estudantes que apesar de todas as adversidades não se deixaram abater e concluíram o curso.

A todos meus mestres, professores e orientadores por me passarem o bem mais precioso, o conhecimento.

As instituições de fomento à pesquisa CNPQ (Conselho Nacional de Pesquisa) e a Petrobras por proporcionarem o apoio financeiro que foi e é fundamental para o desenvolvimento de qualquer pesquisa.

Ao Doutorando em Engenharia Elétrica, Ednardo Moreira Rodrigues, e seu assistente, Alan Batista de Oliveira, aluno de graduação em Engenharia Elétrica, pela adequação do *template* utilizado neste trabalho para que o mesmo ficasse de acordo com as normas da biblioteca da Universidade Federal do Ceará (UFC).

“O homem pode morrer, mas suas ideias são eternas”

(autoral)

RESUMO

O objetivo deste trabalho é gerar variáveis aleatórias discretas com distribuição em lei de potência motivados pelo fato de muitos fenômenos da natureza seguirem essa distribuição. Para alcançarmos tal objetivo, necessitamos entender o *método da inversão* que usaremos para mapear a distribuição binomial na distribuição lei de potência. Como aplicação iremos usar a distribuição gerada por este método para ser usada como distribuição de conectividade em redes complexas.

Palavras-chave: Leis de potência. grafos. *network science*. distribuição de grau.

ABSTRACT

The aim of this work is to generate discrete random variables with power law distribution. Motivated by the fact that a lot of natural phenomena are characterized by this distribution. To reach this task we must understand the *inversion method* that we are going to use to map the binomial distribution into power law distribution. As an application we will use the distribution we generated following this method to be used as a connectivity distribution on network science.

Keywords: Power law. graphs. network science. degree distribution.

LISTA DE FIGURAS

Figura 1	– Estrutura calculada de um <i>cluster</i> de água H_2O com estrutura icosaédrica.	16
Figura 2	– Reação de combustão da lã de aço.	17
Figura 3	– Exemplo de um histogram da distribuição dos acertos de um experimento tipo lei de potência. No contexto de redes complexas o histogram da conectividade média.	23
Figura 4	– Exemplo de um <i>plot</i> feito na escala <i>log – log</i> mostrando a linearidade da função distribuição nesta escala.	24
Figura 5	– Função distribuição de conectividade para uma rede com tamanho 10^6 nós e conectividade $\langle k \rangle = 10$ e $\alpha = 5.0$	27
Figura 6	– Função distribuição de conectividade para uma rede com tamanho 10^6 nós e conectividade $\langle k \rangle = 10$ e $\alpha = 3.0$	28
Figura 7	– Função distribuição de conectividade para uma rede com tamanho 10^6 nós e conectividade $\langle k \rangle = 10$ e $\alpha = 2.0$	29

LISTA DE SÍMBOLOS

κ	conectividade média
k	grau - número de conexões no nó
ρ_i	função densidade de probabilidade
$P_{>}$	Função distribuição acumulada
$B(x; n, P)$	Função distribuição binomial
P_k	Função distribuição lei de potência
γ	Expoente da lei de potência
α	parâmetro de controle do modelo, número real
X	variável aleatória dicreta
N	Número de nós ou o tamanho da rede
$isam$	Número de <i>sample</i>
j	sorteio ou medida
$f_X(x)$	função densidade de probabilidade
F_X	Função distribuição acumulada
k_{ij}	conectividade do nó i dado o sorteio j
i	grau, número de ligações e evento
p	Probabilidade de sucesso
q	Probabilidade de fracasso
m	<i>cutoff</i>
K	<i>cutoff</i>
U	Variável aleatória uniforme

SUMÁRIO

1	INTRODUÇÃO	13
1.1	Introdução	13
1.2	Introdução à Erdos-Renyi	13
2	TRANSIÇÃO DE FASE EM ERDS-RNYI	15
2.1	Material Experimento 1	15
2.2	Experimento 1:	15
2.3	Experimento 2:	17
2.4	Modelo de configuração	18
2.5	Grafos aleatórios livre de escala	19
3	MÉTODO DA INVERSÃO	20
3.1	Variáveis Aleatórias Discretas	20
3.2	Variável Aleatória Contínua	20
3.3	Função Distribuição Acumulada	20
3.4	Distribuição Binomial	21
3.5	Distribuição Lei de Potência	21
3.6	O método da Inversão	24
4	RESULTADOS	27
5	CONCLUSÕES E TRABALHOS FUTUROS	30
	APÊNDICES	32
	APÊNDICE A – A conexão entre as distribuições de Poisson e binomial .	32
	ANEXOS	32

1 INTRODUÇÃO

1.1 Introdução

O foco dessa monografia é gerar variáveis aleatórias discretas com distribuição em lei de potência. Para tal precisaremos entender conceitos que são fundamentais para o entendimento desse método.

Como uma aplicação iremos fazer um paralelo com redes complexas usando a distribuição que geramos a partir do nosso método.

A monografia foi organizada em seções que discutem os conceitos necessários para entendermos o método da inversão. Cada seção foi planejada para ser independente uma da outra, ou seja, não é necessário ler o texto por completo, podemos ir direto as seções que o leitor possui maior interesse.

Fizemos uma breve discussão do conceito de transição de fase que ocorre em modelos de redes complexas. Para um entendimento mais didático escolhi expôr esse conceito através de dois experimentos.

1.2 Introdução à Erdos-Renyi

Grafos aleatórios foram a primeira vez introduzidos por Paul Erdős e Alfred Rényi [1], dois matemáticos Húngaros em 1959, e de maneira independente, no mesmo ano, Edgar Gilbert também cria seu modelo. O Edgar Gilbert era americano e trabalhou nos Laboratórios Bell [2].

Ambos os modelos são semelhantes e hoje os conhecemos por modelo Erdős-Rényi de grafos aleatórios. Um grafo é um conjunto de pontos (nós) e ligações definidos em um espaço de probabilidades. As diferenças entre os modelos são simples. O primeiro modelo $G(n, M)$, onde n são o número de pontos (nós) e M é o número de ligações, simplesmente toma um grafo com n pontos e M ligações. Em $G(n, p)$, o modelo de Gilbert, a situação é a seguinte, cada configuração está inserida no espaço de probabilidades de todas as configurações possíveis de pontos (nós) e ligações equiprováveis. Atribuímos uma probabilidade entre 0 e 1 de uma dessas configurações serem selecionadas. Em $G(n, M)$ a distribuição de ligações segue a distribuição binomial, pois a única coisa que precisamos fazer é combinar essas ligações em seus respectivos pontos, enquanto em $G(n, p)$ segue a distribuição de Poisson para $n \rightarrow \infty$. Um exemplo que representaria essa

diferença seria, a probabilidade que $G \in G(n,p)$ tenha um número par de ligações é um número entre 0 e 1, enquanto que a mesma probabilidade em $G(n,M)$ é exatamente 0 ou 1.

O número de ligações esperados em $G(n,p)$ é $\binom{n}{2}p$, então $G(n,p)$ se comporta de maneira semelhante a $G(n,M)$ quando $M = \binom{n}{2}p$ e $n \rightarrow \infty$.

Em todos os exemplos que trataremos nesta monografia a presença ou a ausência de uma ligação entre dois vértices é independente da ausência ou da presença de qualquer outra ligação, de tal forma que cada ligação deve ser considerada independente com probabilidade p .

Vamos supor que existam N pontos em um determinado grafo, e cada um deles está conectado por uma média λ , então sabemos que $p = \frac{\lambda}{(N-1)}$, que para N grande podemos aproximar por λ/N . O número de ligações de um ponto ou nó qualquer é chamado de grau e denotaremos por k com sua distribuição de probabilidade dada por

$$p_k = \binom{N}{k} p^k (1-p)^{(N-k)} = \frac{\lambda^k e^{-\lambda}}{k!} \quad (1.1)$$

Lembrando que a última igualdade só é válida para o limite de N grande. Esta distribuição é a conhecida distribuição de Poisson. A partir de agora iremos tratar de grafos ou redes que possuem esse tipo de distribuição. [3].

2 TRANSIÇÃO DE FASE EM ERDÖS-RÉNYI

Começaremos explicando o processo de transição de fase com dois experimentos. A seguir temos uma explicação detalhada desses dois experimentos.

2.1 Material Experimento 1

1. Prato
2. Água
3. Pimenta do reino
4. Detergente

2.2 Experimento 1:

Pegue um prato de fundo branco. Deposite neste mesmo prato uma certa quantidade de água de tal forma que metade do volume do prato se preencha. Agora deposite sobre a superfície da água pimenta em pó; de preferência pimenta do reino, para contrastar com o fundo branco do prato. Povolhe pimenta até que as partículas de pimenta se depositem uniformemente sobre a superfície da água. O prato é circular, então pingue uma gota de sabão no centro do prato. Todo o *cluster* de pimenta se quebrará em pedaços.

A água é um complexo de moléculas que são mantidas juntas por atrações do tipo forças de van der Waals ou por pontes de hidrogênio. O átomo de oxigênio é mais eletronegativo do que o átomo de hidrogênio, ou seja, o núcleo do oxigênio atrai os elétrons envolvidos na ligação $O-H$ mais fortemente que o núcleo de hidrogênio.

Essa propriedade é decisiva na polaridade da molécula da água. Pois os elétrons do Oxigênio estão mais próximos do núcleo do que os elétrons dos átomos de Hidrogênio. A molécula da água é uma molécula polar, sendo o seu pólo negativo o Oxigênio e os pólos positivos os Hidrogênios. Sendo a água um composto polar, os pólos positivos atraem os pólos negativos via interação coulombiana. Essa ligação entre o Oxigênio e os Hidrogênios tem nome, ponte de hidrogênio, e o ocorrem entre átomos de Nitrogênio(H_2), Oxigênio(O_2), Flúor(F) ou Nitrogênio(N_2).

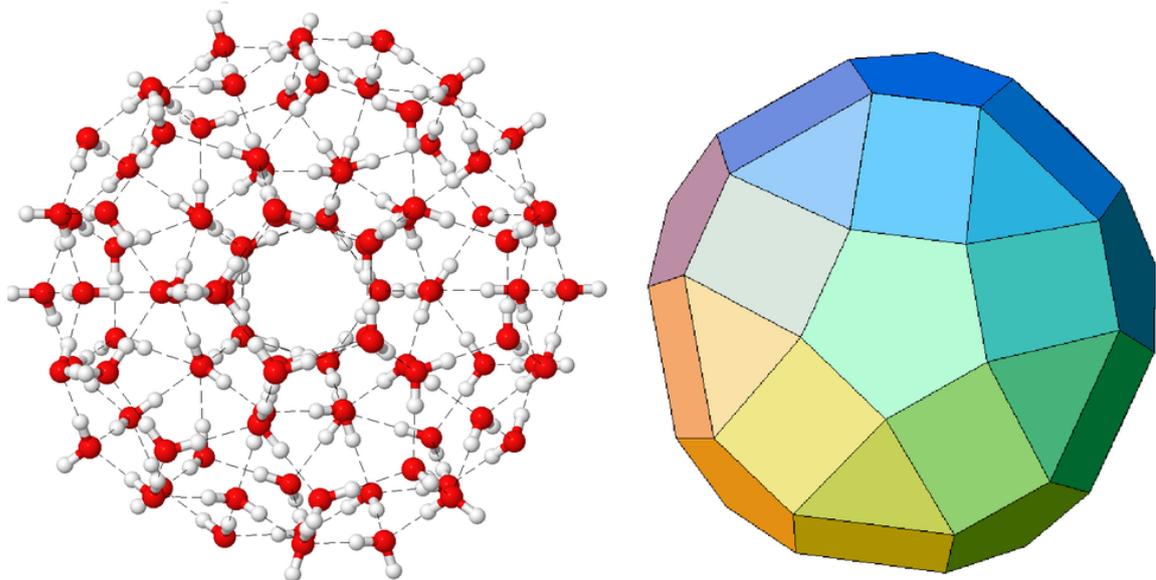
O detergente é considerado um sabão mole. Devido ao fato de possuir um átomo de Potássio(K) ligado a uma cadeia de gordura. O sabão é um sal de ácido carboxílico ($R-COOH$). Aqui também o Oxigênio é responsável por manter a molécula do sabão polar.

Perceba, o que temos é uma mistura de duas moléculas iônicas. Essa diferença de concentração é altamente instável e o composto tende a se estabilizar muito rapidamente. Essa diferença na solubilidade da água causada pela presença do sabão afeta, inclusive a tensão superficial, daí o motivo do sabão ser considerado um surfactante(tensioativo).

O que acabamos de descrever é um processo de transição de fase onde o componente gigante (de pimenta, do reino) se desintegra deixando apenas componentes isoladas. Tudo isso só é possível devido ao fato de a água ser uma molécula de Van der Waals que é um complexo fracamente ligado de átomos ou moléculas mantidas juntas por atrações tais como Van der Waals ou por pontes de hidrogênio é claro, o peso da pimenta não é suficiente para romper essa força. O sabão é um surfactante ou melhor, um tensioativo ($C_{12}H_{25}C_6H_4SO_3Na$) que possui uma longa cadeia de hidrogênio e carbono possuindo no fim dessa cadeia um sal. Quando a gota de sabão se mistura com a água aumentamos a concentração de um dos íons que formam o precipitado, a concentração do outro deve diminuir, para que o produto de solubilidade permaneça constante. Esse efeito permite diminuir a solubilidade de muitos precipitados reduzindo a tensão superficial da água deixando a pimenta baixar.

A seguir temos uma representação de uma *cluster* formado por moléculas de água

Figura 1 – Estrutura calculada de um *cluster* de água H_2O com estrutura icosaédrica.



Fonte: Wikipédia.

2.3 Experimento 2:

Material Experimento 2

1. Bombril (Lã de aço)
2. Fonte de calor (isqueiro, bateria 9V)

Experimento 2: Pegue um pedaço de bombril(lã de aço). Abra toda a folha da lã de aço. Perceba que o que você tem em mãos é um emaranhado de fios da espessura do seu cabelo. Com uma tesoura corte um pedaço na forma de um quadrado de 5 cm de lado. Com uma fonte de calor acenda o centro ou uma das extremidades. Essa energia térmica na forma de calor se propagará na cor laranja e uma reação em cadeia se sucederá. Se a lã de aço estiver com os fios separados o suficiente para o Oxigênio preencher todo o quadrado, essa energia vai percolar, no sentido que atravessará toda a rede de lã de aço, de uma extremidade a outra. O que estamos fazendo é transformando Ferro ($2Fe$) mais Oxigênio (O_2) em óxido de Ferro ($2Fe_2O_3$). Tudo isso só é possível devido a forma da lã de aço. Os fios serem bem finos permitem toda a bucha ser preenchida por Oxigênio e quando há faísca esse Oxigênio alimenta a combustão. Depois de queimada a lã de aço aumenta de massa, devido ao Ferro de baixo teor de Carbono formado como resultado da queima do aço com Oxigênio [4].

Figura 2 – Reação de combustão da lã de aço.



Fonte: Wikipédia.

O que acabei de descrever foram dois processos de transição de fase, onde a principal

característica são as duas fases que os sistemas possuem antes e depois da transição. No modelo de Erdős-Rényi um processo semelhante acontece. Começamos com n pontos que estão inicialmente isolados. Então, começamos a adicionar a esses pontos ligações, tomando quaisquer dois deles e os ligando. Dizemos que dois pontos estão no mesmo componente (*cluster*) se existe uma caminho de ligações conectando todos esses pontos (*path*). Iremos definir o número de ligações no componente gigante por C . Obviamente, se adicionarmos mais ligações a esse componente ele crescerá a tal ponto que todos os pontos (nós) serão partes de um único componente gigante. O que Paul Erdős provou em seu paper de 1960 é que existe uma probabilidade crítica p_c onde o componente gigante aparece. Se $np < 1$ a ordem do tamanho do maior componente nesse regime não excederá um tamanho maior que $O(\log(n))$. Em contrapartida se $np = 1$, então um grafo terá quase que certamente um componente gigante cujo o tamanho será da ordem de $n^{2/3}$ e por fim, se $np > 1$, então o sistema terá certamente um único componente gigante contendo uma fração positiva dos pontos (nós) do sistema. Perceba que o processo de formação do componente gigante em Erdős-Rényi é de fato um processo de transição de fase e digo além, um processo de transição de fase de segunda ordem, contínuo. Agora faremos uma extensão desse modelo [5][6].

2.4 Modelo de configuração

Como definido anteriormente o grau de um ponto (nó) em uma rede é o número de ligações incidentes (i.e., conectados) a este ponto(nó). Seja p_k a fração de pontos na *network* que tem grau k . De maneira equivalente poderíamos definir p_k como sendo a probabilidade de um nó ser escolhido uniformemente aleatório e tal nó tenha o grau k . Nos modelos de grafos aleatórios estudados por Erdős e Rényi, cada ligação está presente ou ausente com igual probabilidade, e portanto a distribuição de grau desses modelos segue a binomial ou a distribuição de Poisson no limite $n \rightarrow \infty$. As redes reais apresentam um comportamento diferente dos grafos aleatórios no que diz respeito a distribuição de grau dessas redes. Por exemplo, a internet ou WWW é muito mais densa. Os nós possuem em média um número de ligações maior do que os nós em Erdős e Rényi [5][6].

2.5 Grafos aleatórios livre de escala

Um grafo livre de escala é um grafo tendo uma certa distribuição de grau como definido anteriormente. De maneira mais precisa a probabilidade que um nó tenha k ligações

$$P_k = ck^\gamma, k = m, m + 1, \dots, K, \quad (2.1)$$

onde $c \approx (\gamma - 1)m^{\gamma-1}$ é um fator de normalização, m e K são respectivamente os *cutoffs* superiores e inferiores da distribuição [5][6].

3 MÉTODO DA INVERSÃO

Nesta seção iremos discutir os conceitos de variável aleatória, função distribuição, função distribuição binomial e lei de potência. Com esses conceitos bem fundamentados introduciremos o método da transformação inversa que é no final das contas tema principal deste trabalho.

3.1 Variáveis Aleatórias Discretas

Uma variável aleatória discreta é uma variável aleatória tomando somente valores inteiros não-negativos. Na teoria da probabilidade, uma variável aleatória discreta é uma variável aleatória que toma valores com probabilidade 1 em um conjunto contável de pontos dado. Contanto que haja uma correspondência um a um ou uma bijeção entre os conjuntos contáveis e os inteiros não negativos[1].

A distribuição de uma variável aleatória discreta X é determinada pelo vetor probabilidade p_0, p_1, \dots :

$$P(X = i) = p_i (i = 0, 1, 2, \dots). \quad (3.1)$$

3.2 Variável Aleatória Contínua

Na teoria da probabilidade uma variável aleatória é entendida como uma função medida definida em um espaço de probabilidade cujo os resultados são tipicamente números reais. Uma variável aleatória é tipicamente representada por X .

3.3 Função Distribuição Acumulada

No contexto de teoria da probabilidade e estatística, a *função distribuição acumulada (FDA)* de uma variável real aleatória X , medida em x , é a probabilidade que X tomará um valor menor ou igual x .

Se X é uma variável aleatória puramente discreta, então teremos x_1, x_2, \dots com probabilidade $p_i = P(x_i)$, e a FDA será descontínua nos pontos x_i :

$$F_X(x) = P(X \leq x) = \sum_{x_i \leq x} P(X = x_i) = \sum_{x_i \leq x} p(x_i) \quad (3.2)$$

Se a função distribuição acumulada de uma variável aleatória é contínua, então existe a função $f_X(x)$ tal que

$$F_X(b) - F_X(a) = P(a < X \leq b) = \int_a^b f_X(x) dx \quad (3.3)$$

para todo número real a e b . A função f_X é igual a derivada de F_X *almost everywhere*, e é chamada a função densidade de probabilidade da distribuição de X .

3.4 Distribuição Binomial

A distribuição binomial pode ser pensada simplesmente como a probabilidade de sucesso ou fracasso num experimento que é repetido um número grande de vezes. A distribuição binomial é um tipo de distribuição associada a eventos que possuem dois possíveis resultados de onde vem o prefixo bi. Como exemplo podemos citar o lançamento de uma moeda e seus dois possíveis resultados, a saber cara ou coroa. Se jogarmos a moeda um número muito grande de vezes, a distribuição dos possíveis resultados será uma distribuição binomial.

A fórmula da distribuição binomial pode ser escrita na forma

$$B(x; n, P) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x} \quad (3.4)$$

onde:

B - probabilidade binomial

x - total de eventos com sucesso

P - probabilidade de um sucesso em uma única tentativa

n - número de tentativas

Aqui temos que ressaltar o fato de p ser a probabilidade de obter sucesso e q a probabilidade de fracasso com $p + q = 1$.

No contexto de redes, podemos interpretar o fato de um nó está conectado a outro nó como uma probabilidade de sucesso ou fracasso e entendermos a distribuição de conectividade como uma distribuição Binomial. Onde a probabilidade de cada nó está ligado a outro nó é identicamente independentemente distribuída (*i.i.d.*) refletindo o fato de $p + q = 1$.

3.5 Distribuição Lei de Potência

O objetivo dessa seção é expor o conceito de lei de potência dentro do contexto de redes complexas motivados pelo trabalho (power law parets...). Dentro do contexto de redes

complexas, seja uma rede onde a distribuição de conectividade segue uma distribuição tipo lei de potência. Nesse caso o que teremos é um grande número de nós com uma conectividade média baixa e uma pequena quantidade de pontos com a conectividade média alta. Sendo a distribuição altamente *right-skewed*, no sentido que a grande maioria dos nós possuem conectividade média baixa e uma pequena minoria com conectividade média alta.

Façamos a seguinte observação (Auerbach [1]). Seja $p(x)dx$ a fração de nós com conectividade entre x e $x + dx$. Se o histograma desses nós é uma linha reta na escala $\log - \log$, então $\ln p(x) = -\alpha \ln x + c$, onde α e c são constantes. Tomando a exponencial de ambos os lados,

$$p(x) = Cx^{-\gamma}, \quad (3.5)$$

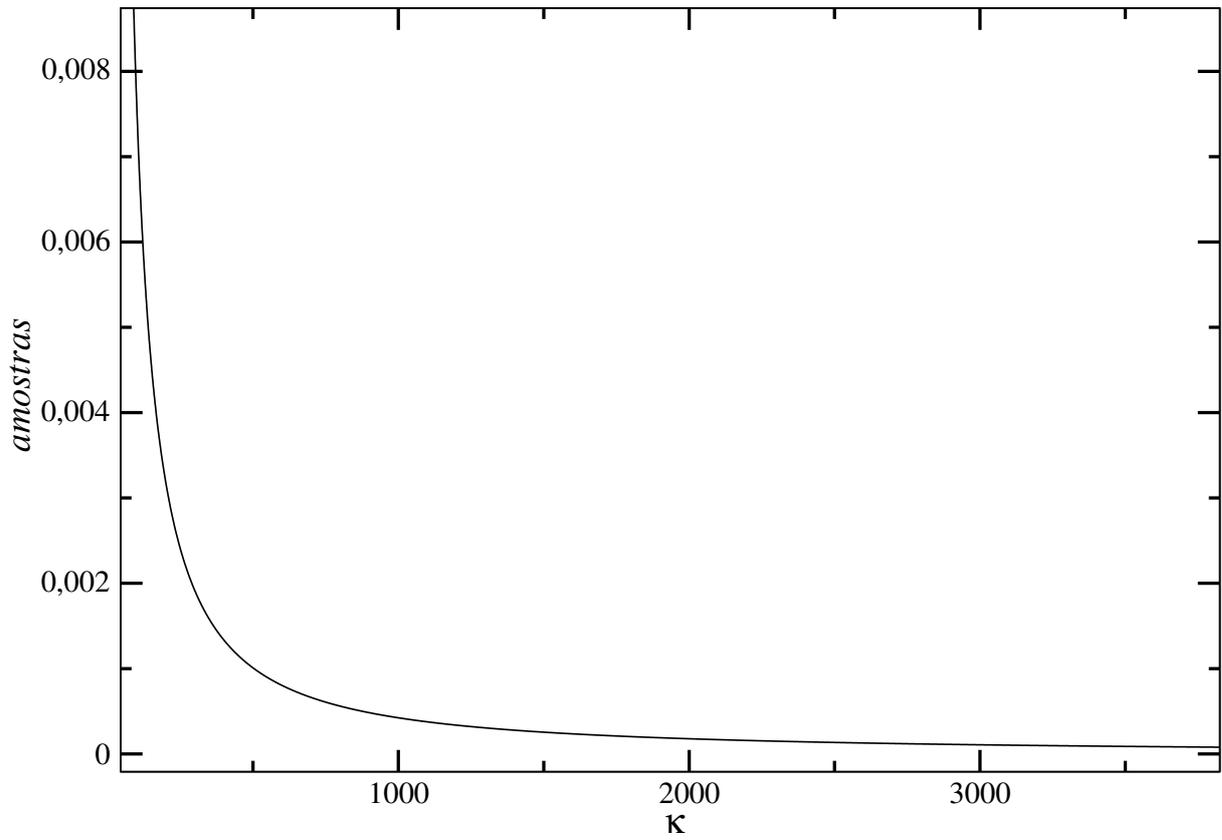
com $C = \exp(c)$.

A distribuição na forma acima é dita lei de potência e a constante γ é chamada o *expoente* da lei de potência.

Distribuição tipo lei de potência ocorrem em muitos fenômenos na natureza. Além da distribuição de conectividade em redes complexas, as crateras na lua(4), labaredas solares(3), o número de artigos que os cientistas escrevem(10), número de citações recebidos por paper(11), o número de acessos recebidos numa página na internet(12) e a lista continua mostrando muitos exemplos de fenômenos que ocorrem sob a forma dessa distribuição.

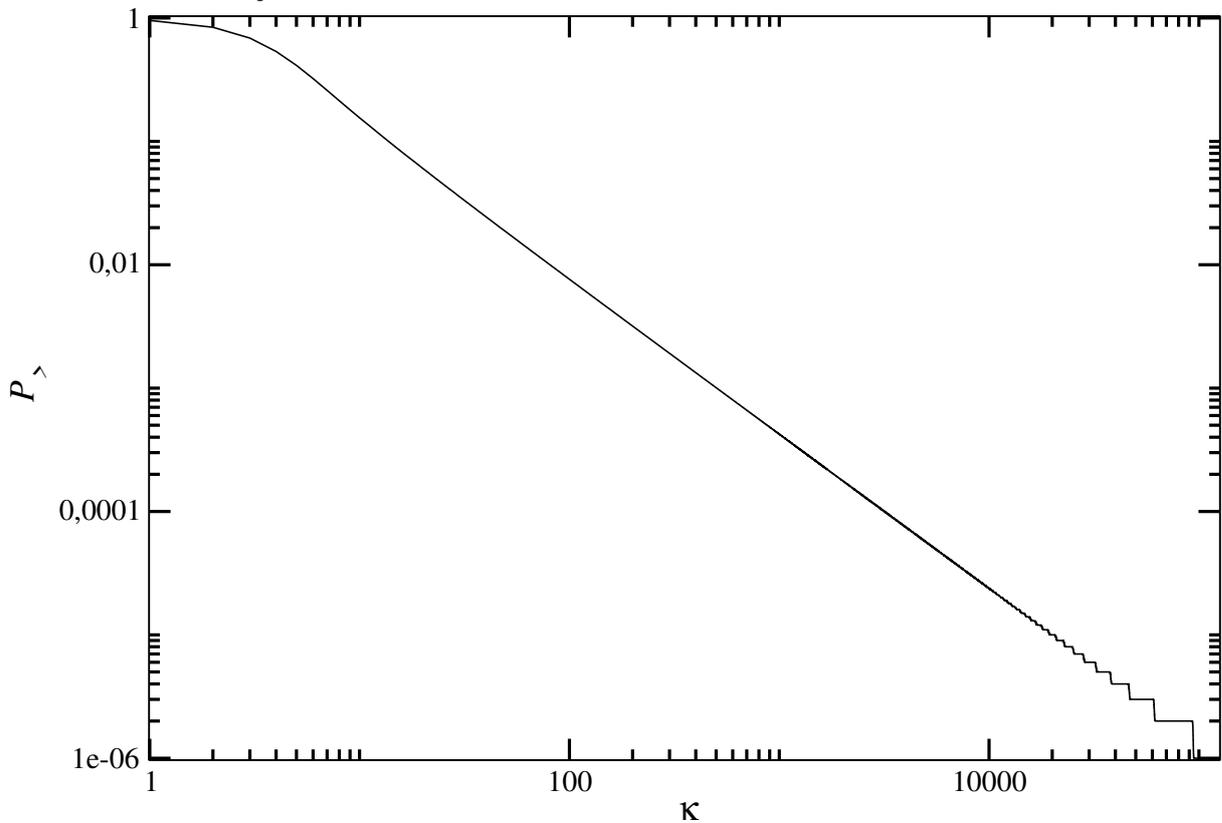
A seguir temos dois exemplos com dados gerados a partir de um programa que criamos. Na primeira figura o que temos é simplesmente um histograma do número de acertos ou ligações médias e no segundo o que temos o mesmo histograma na escala $\log - \log$.

Figura 3 – Exemplo de um histogram da distribuição dos acertos de um experimento tipo lei de potência. No contexto de redes complexas o histogram da conectividade média.



Fonte: Fonte autoral.

Figura 4 – Exemplo de um *plot* feito na escala *log – log* mostrando a linearidade da função distribuição nesta escala.



Fonte: Fonte autoral.

3.6 O método da Inversão

No **método da inversão**, nós geramos uma variável aleatória uniforme U em $[0, 1]$ e obtemos X através de uma transformação monotônica de U que é da forma $P(X = i) = p_i$. Se nós definimos X por

$$F(X - 1) = \sum_{i < X} p_i < U \leq \sum_{i \leq X} p_i = F(X) \quad (3.6)$$

então $P(X = i) = F(i) - F(i - 1) = p_i$. Este procedimento é comparável ao método da inversão para variáveis contínuas [10].

Exemplo: Seja i uma variável aleatória discreta da forma

$$i = \lfloor N \times R^\alpha \rfloor \quad (3.7)$$

onde α é um número real positivo (\mathbb{R}^*), N uma constante e R um número aleatório real (\mathbb{R}) em $[0, 1]$. É preciso notar que o caso $\alpha = 1.0$ é o modelo Erdos and Rényi. Perceba que $\lfloor N \times R^\alpha \rfloor \equiv$ maior inteiro $\leq N \times R^\alpha$. Se $i \leq N \times R^\alpha < i + 1$ ou $(\frac{i}{N})^{\frac{1}{\alpha}} \leq R < (\frac{i+1}{N})^{\frac{1}{\alpha}}$, podemos escrever a

probabilidade ρ_i de obter sucesso no evento i da forma

$$\rho_i = \left(\frac{i+1}{N}\right)^{\frac{1}{\alpha}} - \left(\frac{i}{N}\right)^{\frac{1}{\alpha}}. \quad (3.8)$$

Podemos escrever o primeiro termo da seguinte forma

$$\left(\frac{i+1}{N}\right)^{\frac{1}{\alpha}} = \left(\frac{i}{N} \left(1 + \frac{1}{i}\right)\right)^{\frac{1}{\alpha}} = \left(\frac{i}{N}\right)^{\frac{1}{\alpha}} \left(1 + \frac{1}{i}\right)^{\frac{1}{\alpha}}. \quad (3.9)$$

Expandindo $\left(\frac{i+1}{N}\right)^{\frac{1}{\alpha}}$ em série de Taylor

$$\left(\frac{i+1}{N}\right)^{\frac{1}{\alpha}} \simeq \left(\frac{i}{N}\right)^{\frac{1}{\alpha}} \left(1 + \frac{1}{\alpha} \frac{1}{i} + \dots\right) \quad (3.10)$$

Ou seja, para i grande o suficiente

$$\rho_i \simeq \frac{i^{\frac{1-\alpha}{\alpha}}}{\alpha N^{\frac{1}{\alpha}}}. \quad (3.11)$$

Cada Π_i é um elemento que compõe o vetor probabilidade $P(X = i) = \rho_i$. Para $i \gg 1$ podemos tratar como uma distribuição contínua. Note que $\rho_i = \frac{dF}{di}$.

Então a probabilidade de selecionar um ponto (nó) k_i é aproximadamente

$$k_i = \rho_i \times M \quad (3.12)$$

onde M é o número total de ligações dessa rede. É preciso notar que, para $\alpha > 1$ em média, k_i decresce quando i cresce.

O que precisamos fazer agora é encontrar a distribuição de probabilidade de k ou $P_k(k)$ a partir da distribuição dos i 's. Tratando k e i como variáveis contínuas e notando que existe o mapeamento de uma distribuição na outra, podemos escrever

$$P_k(k)dk = P_i(i)di \quad (3.13)$$

Devido ao fato de $1 \leq i < N + 1$ e todos os P_i 's serem igualmente prováveis

$$\int_1^{N+1} P_i(i)di = 1 \quad (3.14)$$

de maneira que

$$P_i(k) = \frac{1}{N}. \quad (3.15)$$

Logo,

$$P_k(k) = -\left|\frac{1}{N} \frac{d_i(i)}{d_k(k)}\right|. \quad (3.16)$$

Lembrando que

$$k(i) = \frac{Mi^{\frac{1-\alpha}{\alpha}}}{\alpha N^{\frac{1}{\alpha}}}. \quad (3.17)$$

Invertendo i

$$i = \left(\frac{\alpha N^{\frac{1}{\alpha}} k}{M} \right)^{\frac{\alpha}{1-\alpha}}. \quad (3.18)$$

Derivando i em relação a k

$$\frac{d(i)}{d(k)} = \left(\frac{\alpha}{M} \right)^{\frac{\alpha}{1-\alpha}} N^{\frac{1}{1-\alpha}} \frac{\alpha}{1-\alpha} k^{-\frac{2\alpha-1}{\alpha-1}} \quad (3.19)$$

Finalmente,

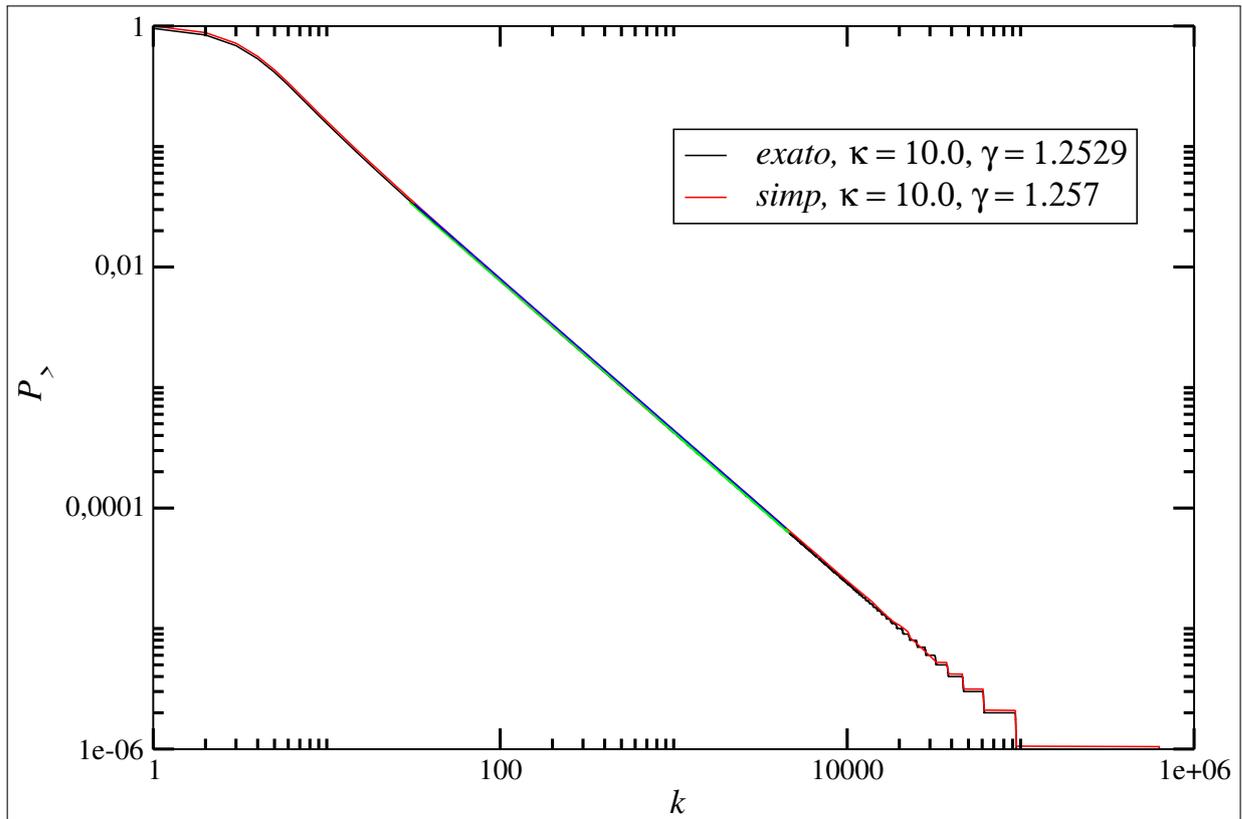
$$P_k(k) = \left(\frac{N}{M} \right)^{\frac{\alpha}{1-\alpha}} \frac{\alpha^{\frac{1}{1-\alpha}}}{1-\alpha} k^{1-\frac{2\alpha-1}{\alpha-1}} \quad (3.20)$$

que é a tão esperada distribuição de probabilidade para os eventos k_i .

4 RESULTADOS

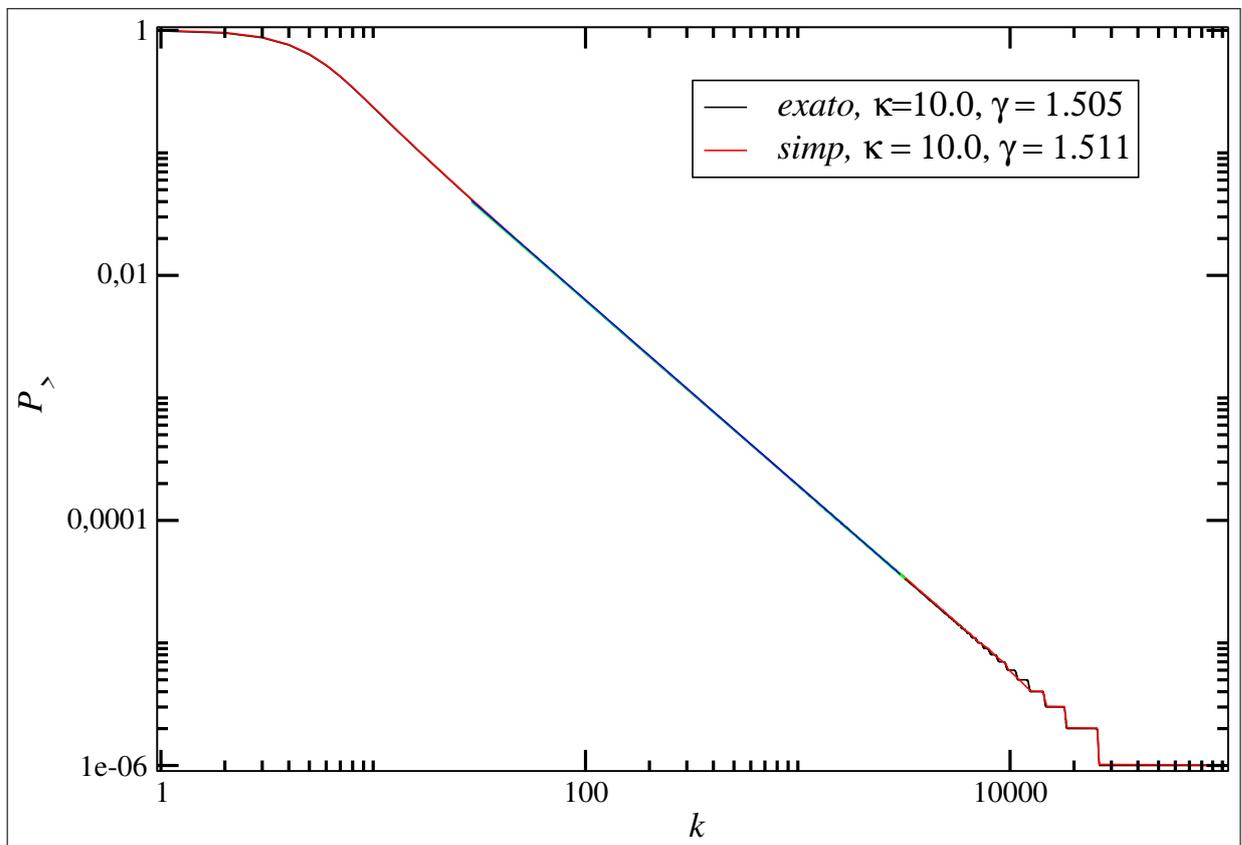
Construímos um programa em $C++$ que mapeada as distribuições analiticamente resolvendo as integrais. Em um outro programa geramos dados apartir de um gerador de números aleatórios e testamos esses dados numericamente. A seguir *plotamos* as distribuições obtidas comparando os resultados numérico e anflítico para uma rede com 10^6 pontos e $\alpha = 5.0$.

Figura 5 – Função distribuição de conectividade para uma rede com tamanho 10^6 nós e conectividade $\langle k \rangle = 10$ e $\alpha = 5.0$.



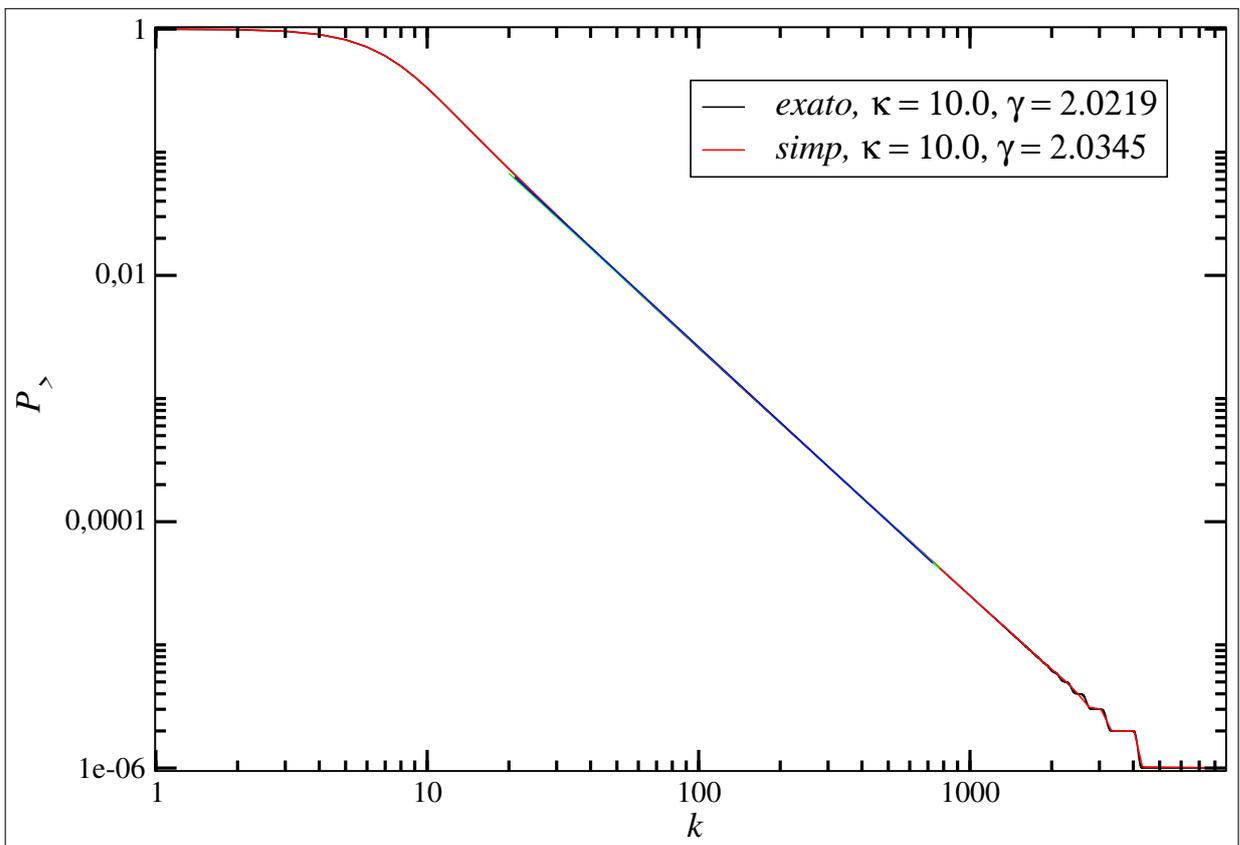
Fonte: elaborado pelo autor (2019).

Figura 6 – Função distribuição de conectividade para uma rede com tamanho 10^6 nós e conectividade $\langle k \rangle = 10$ e $\alpha = 3.0$.



Fonte: elaborado pelo autor (2019).

Figura 7 – Função distribuição de conectividade para uma rede com tamanho 10^6 nós e conectividade $\langle k \rangle = 10$ e $\alpha = 2.0$.



Fonte: elaborado pelo autor (2019).

5 CONCLUSÕES E TRABALHOS FUTUROS

Construímos um programa em $C++$ que gera números aleatórios. A partir dessas sequências contamos o número de sucessos em nosso experimento e achamos a função distribuição para esses dados. Com a função distribuição em mãos, lançamos mão do método da inversão para conseguirmos a distribuição acumulada. No método da inversão escolhemos como função alvo a distribuição em forma de lei potência pelo fato de tal distribuição aparecer em diversos fenômenos da natureza como exemplificado anteriormente.

Como perspectivas futuras podemos utilizar o nosso método para gerar distribuições genuinamente na forma de lei de potência para testar dados experimentais e estudar o surgimento do componente gigante para caracterizar esse processo de transição de fase.

Além de podermos usar nossos dados para alimentar a máquina e com algum algoritmo de aprendizagem de máquina podermos caracterizar e elucidar o comportamento de redes complexas.

Referências

- [1] Erdős, P.; Rényi, A. (1959). "On Random Graphs. I". *Publicationes Mathematicae*. 6: 290–297.
- [2] Gilbert, E.N. (1959). "Random Graphs". *Annals of Mathematical Statistics*. 30 (4): 1141–1144. doi:10.1214/aoms/1177706098.
- [3] Bollobás, B. (2001). *Random Graphs* (2nd ed.). Cambridge University Press. ISBN 0-521-79722-5.
- [4] James Gordon and Katherine Chancey, 2005, *Steel Wool and Oxygen: A Look at Kinetics*, doi: 10.1021/ed082p1065.
- [5] Reuven Cohen, Shlomo Havlin, (2010), *Complex Networks: Structure, Robustness and Function*. ISBN: 9780521841566.
- [6] Mark Newman, (2010), *Networks an introduction*. ISBN: 9780199206650.
- [7] M. E. J. Newman and R. M. Ziff, (2001), Fast Monte Carlo algorithm for site or bond percolation, *Phys. Rev. E*. DOI:<https://doi.org/10.1103/PhysRevE.64.016706>.
- [8] M. E. J. Newman G. T. Barkema(2001).*Monte Carlo Methods in Statistical Physics*. Oxford. ISBN 019 851796 3 (Hbk).
- [9] Ginestra Bianconi, (2009), Entropy of network ensembles, *Phys. Rev. E*. DOI:<https://doi.org/10.1103/PhysRevE.79.036114>.
- [10] Luc Devroye (1986). "Section 2.2. Inversion by numerical solution of $F(X) = U$ ". *Non-Uniform Random Variate Generation*. New York: Springer-Verlag.

APÊNDICE A – A CONEXÃO ENTRE AS DISTRIBUIÇÕES DE POISSON E BINOMIAL

A distribuição de Poisson é de fato o caso limite da distribuição Binomial, quando o número de tentativas n cresce muito e p , a probabilidade de sucesso, é pequena.

Considere a probabilidade de obtermos sucesso em um evento k em n tentativas. Definido da forma a baixo.

$$P(k) = \binom{n}{k} p^k q^{n-k} \quad (\text{A.1})$$

Denotaremos o valor esperado da distribuição binomial, np , por λ . Note que $p = \frac{\lambda}{n}$ desde que $q = 1 - p$,

$$q = 1 - \frac{\lambda}{n} \quad (\text{A.2})$$

Agora podemos usar isso para escrever $P(k)$ em termos de λ , n e k , temos

$$P(k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \quad (\text{A.3})$$

Abrindo o binômio de Newton tomado k vezes

$$P(k) = \frac{n(n-1)(n-2)\dots(n-k+1)}{k!} \cdot \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} \quad (\text{A.4})$$

Note que existem exatamente k fatores no denominador da primeira fração. Vamos trocar de posição os denominadores da primeira e da segunda fração, separando os n^k por todos os elementos do numerador da primeira fração.

$$P(k) = \frac{n}{n} \cdot \frac{n-1}{n} \dots \frac{n-k+1}{n} \cdot \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^{n-k} \quad (\text{A.5})$$

Agora vamos separar o último fator em dois pedaços

$$P(k) = \frac{n}{n} \cdot \frac{n-1}{n} \dots \frac{n-k+1}{n} \cdot \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \quad (\text{A.6})$$

Tomando o limite $n \rightarrow \infty$ e mantendo k e λ fixos, as primeiras k frações tenderão à 1, assim como o último termo. O segundo termo é um limite conhecido da exponencial, $e^{-\lambda}$, os outros termos permanecem inalterados, pois eles não dependem de n . De tal forma que

$$\lim_{n \rightarrow \infty} P(k) = \frac{e^{-\lambda} \lambda^k}{k!} \quad (\text{A.7})$$

Como queríamos demonstrar.