Redes Complexas
Aula 2

Aula passada

- Logística e regras
- Redes por todos os lados
- Redes Complexas

Aula de hoje

- Representando redes
- Falando sobre redes
- Grau, distância, clusterização

Rede

- Abstração que permite **codificar** relacionamentos entre **pares** de objetos

objetos → vértices do grafo

relacionamentos → arestas do grafo

Exemplos?

Do que se trata Redes Complexas?

- Entender como e porque as coisas se conectam e as consequências desta conectividade

"Coisas que se conectam" → **Redes**

"Como, por que, e consequências" → **Complexo**

- ***Estrutura*** assume papel central
  - o necessária para compreender fenômenos que ainda não explicamos
- A rede não é complexa!

Objetos e Relacionamentos

- Sobre objetos
  - idênticos, diferentes tipos, possuir estado (rótulos)
  - ex. pessoas, homens e mulheres, nascimento
- Sobre relacionamentos
  - simétricos, assimétricos, diferentes intensidades (pesos), negativos
  - múltiplos relacionamentos na mesma rede
  - ex. amizade, colaboração, interação, confiança, ...

Rede captura estrutura do sistema

Classe de Redes

- **Redes sociais**: relacionamento entre pessoas ou grupo de pessoas
- **Redes de informação** (de conhecimento): codificam associação entre informação
- **Redes tecnológicas**: construída pelo homen geralmente para distribuir *commodities*
- **Redes biológicas**: codifica relacionamentos em sistemas biológicos

Classificação para referência

- *Twitter*: rede social ou rede de informação?

Como Falar sobre Redes?

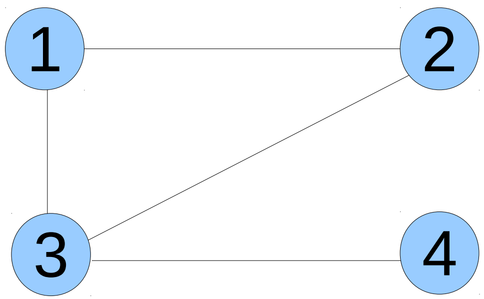
- Desenho da rede: “figura vale por mil palavras”
- Não funciona para redes grandes

Descrever estrutura da rede!

- **Matriz de adjacência**
- Matriz *n* x *n* (*n* é número de vértices)
  - $a_{ij} = 1$, se existe aresta entre vértices *i* e *j*
  - $a_{ij} = 0$, caso contrário
- Codifica todos os relacionamentos da rede
- Generaliza com pesos, pode ser assimétrica

Matriz de Adjacência

- Exemplo



	1	2	3	4
1	0	1	1	0
2	1	0	1	0
3	1	1	0	1
4	0	0	1	0

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

?

Matriz de Adjacência

- Problema em descrever a estrutura com matriz de adjacência?
- Possui toda a informação mas pouco intuitiva
- **Matriz é o DNA da rede**
  - Como descrever uma pessoa? Através do seu DNA?

Precisamos de resumos da estrutura!

- Resumos que ajudem a entender a estrutura de forma intuitiva

Características de Redes

- Caracterísiticas estruturais são resumos da estrutura
  - ex: tamanho, densidade, graus, distâncias, clusterização, centralidade, homofilia, etc
- Dão ideia geral da estrutura da rede

- Quais características devem ser avaliadas? Quais são importantes?
- Depende do propósito!
- Como gens que formam o DNA, estamos apenas começando a enteder seu significado

Característica Importante

- O que faz uma característica ser importante?

1) Prever (determinar) comportamento geral de algum processo
  - independente de outras características

2) Influência sobre diferentes processos

- Exemplo: distribuição de grau
  - determina comportamento de passeios aleatórios
  - influencia outros processos (epidemia)
- Não conhecemos muitas características fundamentalmente importantes
  - a serem descobertas (como gens importantes)

Vértices e Arestas

- Número de vértices de um grafo
  - n = |V|

cardinalidade
do conjunto

- Número de arestas de um grafo
  - m = |E|
- Dado G = (V, E), qual é o maior número de arestas de G?
  - número de pares não ordenados em um conjunto de *n* = |V| objetos

$$\binom{n}{2} = \frac{n(n-1)}{2} \leq n^2$$

- Densidade: fração de arestas que o grafo possui

$$\rho = \frac{m}{\binom{n}{2}} = \frac{2m}{n(n-1)}$$

Grau

- Número de arestas (relacionamentos) incidente em um vértice
  - g(u) : grau do vértice u

- Como falar sobre o grau dos vértices da rede?
- Como falar de uma característica sobre um conjunto de objetos?
- Média, desvio padrão: um número
- Distribuição empírica: todos os graus – comportamento geral

Grau Médio

- Grau médio do grafo considerando todos seus vértices

$$\bar{g} = 1/n \sum_{u \in V} g(u)$$

Grau do vértice u

- Também pode ser calculado diretamente

$$\sum_{u \in V} g(u) = 2m$$

- Cada aresta tem duas pontas!
- Logo temos $\bar{g} = 2m/n$

Distribuição do Grau

- Distribuição empírica do grau dos vértices
  - frequência relativa do grau

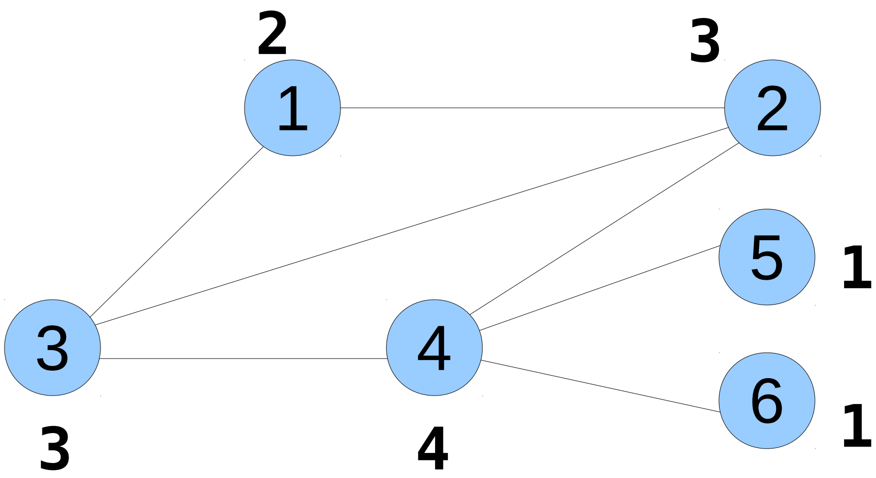
$$f_k = \frac{\text{Número de vértices com grau k}}{\text{Número total de vértices}}$$

fração de vértices com grau k

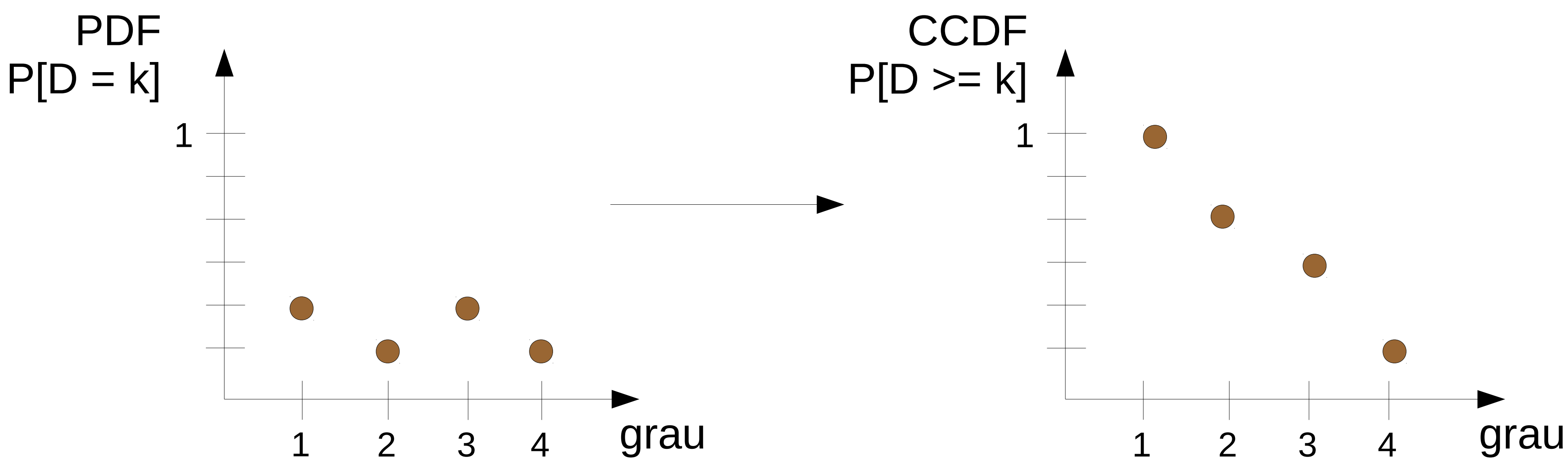
- CCDF empírica (*Complementary Cumulative Distribution Function*)
  - fração de vértices com grau >= k

$$F_k = 1 - \sum_{i=0}^{k-1} f_k$$

Exemplo de Distribuição



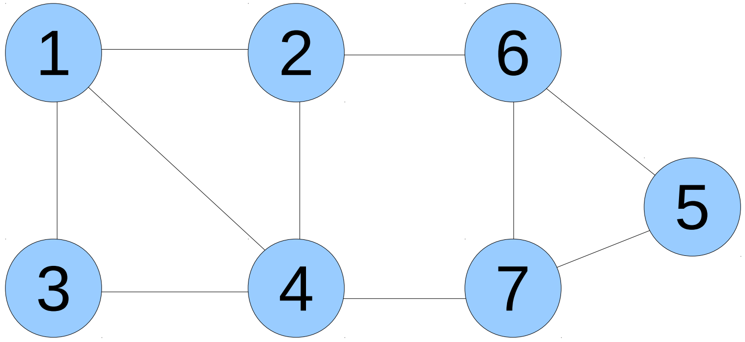
- Distribuição empírica de grau?



Figueiredo – 2015

Distância

- Comprimento do **menor** caminho entre dois vértices
- Função d(u,v), onde u e v são vértices
  - não definido quando não há caminho



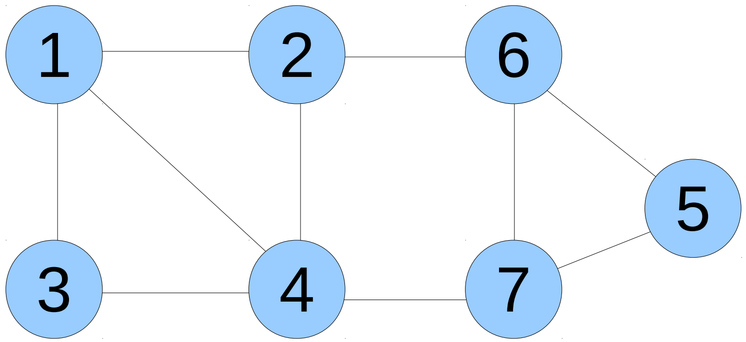
- Exemplo
  - d(6, 3) = ?
  - d(7, 1) = ?

- **Importante**: muitos caminhos podem realizar a distância entre dois vértices

Distância Média e Diâmetro

- Distância média do grafo
  - média entre todos os pares de vértices

$$\bar{d} = \frac{\sum_{u,v \in V} d(u,v)}{\binom{n}{2}}$$



- **Excentricidade**: maior distância de um vértice a todos os outros

$$e(u) = max_{v \in V} d(u,v)$$

- **Diâmetro**: maior distância entre dois vértices da rede

$$r = max_{u,v \in V} d(u,v)$$

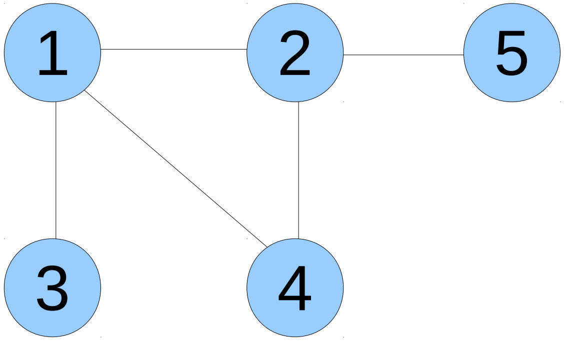
Distribuição da Distância

- Distribuição empírica da distância entre os vértices do grafo
  - frequência relativa das distâncias

$$f_D(d) = \frac{n(d)}{\binom{n}{2}}$$

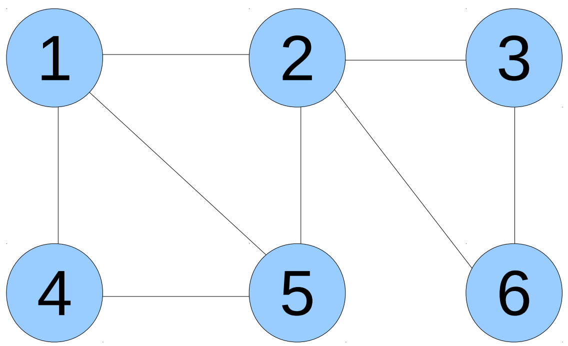
Número de pares não ordenados de vértices que tem distância d

- Exemplo



Clusterização

- Medida da propenção de triângulos se formarem na rede
  - triângulo: ciclo de comprimento três
  - quase-triângulo: caminho de comprimento dois



- Duas métricas na literatura
  - métrica local: cada vértice tem o seu
  - métrica global: uma para a rede

Clusterização – I

- Fração de aresta entre vizinhos
  - prob. de dois vizinhos também serem vizinhos
- Definida para cada vértice da rede

$$C_i = \frac{E_i}{\binom{d_i}{2}}$$

E_i : # de arestas entre os vizinhos de i

d_i : grau do vértice i

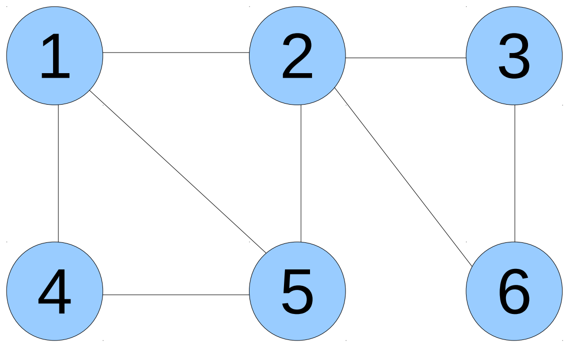
$$C_i = \frac{2E_i}{d_i(d_i - 1)}$$

- Clusterização do grafo
  - média da clusterização dos vértices

$$C = 1/n \sum_{v \in V} C_i$$

Calculando Clusterização

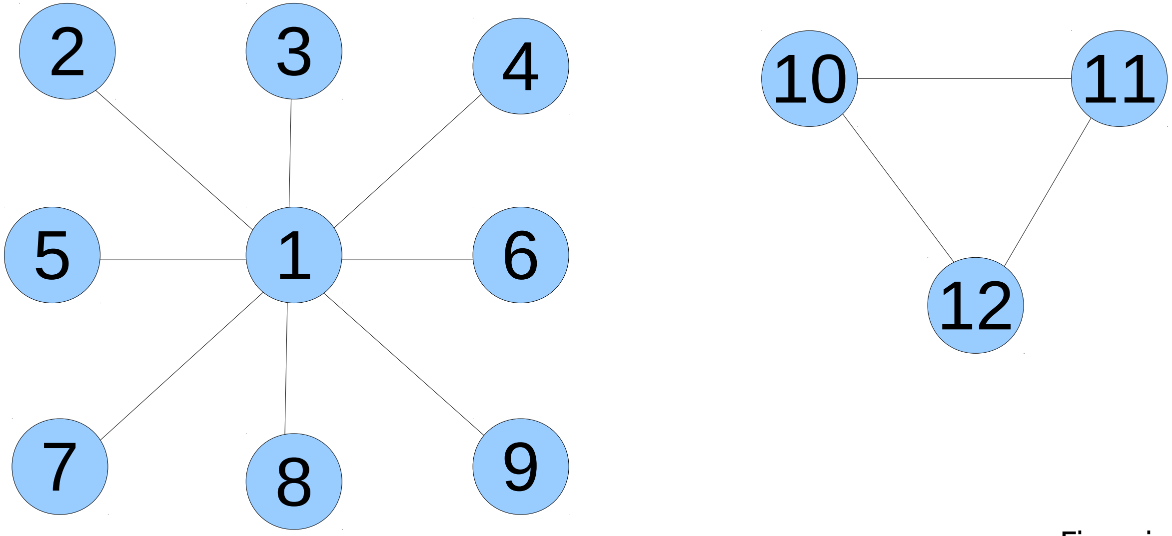
- Exemplo



- Quanto vale C?

Clusterização – I

- **Problemas?**
- Não é a fração de arestas entre vizinhos do grafo (como um todo)
- Favorece vértices com menos vizinhos
  - denominador é ~ d_i^2
- Clusterização para grau 0 e 1?
- Exemplo?



Figueiredo – 2015

Clusterização – II

- Fração entre o número de triângulos e o número de pseudo-triângulos no grafo como um todo
  - pseudo-triângulo: tripla conectada

$$C' = \frac{3 \text{ x número de triângulos}}{\text{número de triplas conectadas}}$$

Cada triângulo dá origem a três triplas conectadas

- Tripla é não-ordenada (“a,b,c” igual a “c,b,a”)
- Métrica global
  - não é média dos vértices
- Não tem problemas com vértices de grau 0 e 1
  - mais bem comportada a anomalias
- Valor pode ser muito diferente da outra definição

Clusterização – II

- Definição equivalente (mesmo valor)
- Pseudo-triângulo: caminho de comprimento dois
- Para cada vértice, número de vizinhos dos meus vizinhos = número de caminhos de comprimento dois
  - Mais fácil de calcular computacionalmente

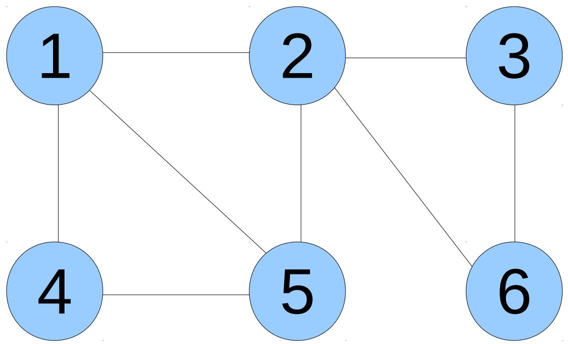
$$C' = \frac{6 \text{ x número de triângulos}}{\text{número de caminhos de comprimento 2}}$$

Cada triângulo dá origem a seis caminhos de comprimento 2

- Igual a métrica anterior, mas agora a tripla é ordenada: (a,b,c) != (c,b,a) logo, conta duas vezes

Calculando Clusterização

- Exemplo



- Quanto vale C?
- Quanto vale C'?