Intro to Complex and Social Networks

Argimiro Arratia & R. Ferrer-i-Cancho

Universitat Politècnica de Catalunya

Complex and Social Networks (2023-2024)
Master in Innovation and Research in Informatics (MIRI)

Instructors

- Ramon Ferrer i Cancho
  - `rferrericancho@cs.upc.edu`
  - Omega S124, 93 413 4028
- Argimiro Arratia
  - `argimiro@cs.upc.edu`
  - Omega 323,

Website

Please go to http://www.cs.upc.edu/~CSN for all course's material, schedule, lab work, etc.

Class Logistics

- Wednesday, 10:00 – 12:00, A6 203
  - Theory lectures.
- Thursday, 10:00 – 12:00, every two weeks, C6 S301.
  - Guided lab activities; expected to be complemented with an average estimate of 4-6 additional hours per session of autonomous lab activities.
  - Lab sessions will require handing in a short written report; these count towards the evaluation of the course.
  - Start on the 7th of September

Lab work - important rules

- Lab reports in teams of 2, submission by one member.
- Work with a different partner each lab.
- Do not exchange information other than general ideas with others: that will be considered plagiarism

Evaluation

There will be no exam in this course. Grading is done entirely through reports on various tasks throughout the course.

- You are expected to hand in 7 lab work reports
  - The best 5 count for 50% of the final grade
  - Lab reports not handed in penalize, so please complete all of them
- You have to do a final course project
  - Project ideas given by instructors towards the end of the course
  - Students pick a project or can suggest their own
  - 50% of the final grade

Contents

As planned today – may go through unpredictable changes

1. Characterization of networks (*how can we describe them*)
   - Lectures 1–7
   - Includes: small-world, degree distribution, finding communities, and other advanced metrics
2. Dynamics of growing networks (*how do networks grow*)
   - Lectures 8–9
   - Includes: random growth, preferential attachment, and other growth models
3. Processing networks and processes on networks (*how can we process large networks and how are processes over networks affected by their topology*)
   - Lectures 10–13
   - Includes: sampling, epidemic models of diffusion, rumor spreading, search, percolation, etc.

So, let's start! Today, we'll see:

1. Examples of real networks
2. What do real networks look like?
    - real networks exhibit small diameter
        - .. and so does the Erdös-Rényi or random model
    - real networks have high clustering coefficient
        - .. and so does the Watts-Strogatz model
    - real networks' degree distribution follows a power-law
        - .. and so does the Barabasi-Albert or preferential attachment model

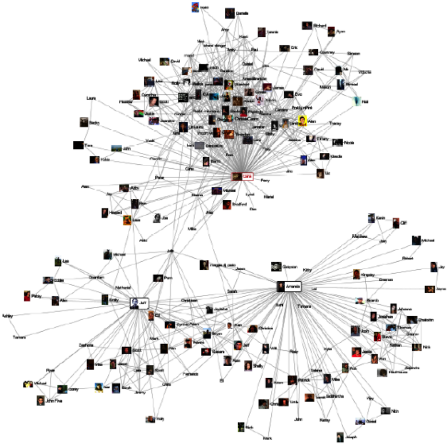
Examples of real networks

- Social networks
- Information networks
- Technological networks
- Biological networks
- Financial networks

Social networks

Links denote social "interactions"

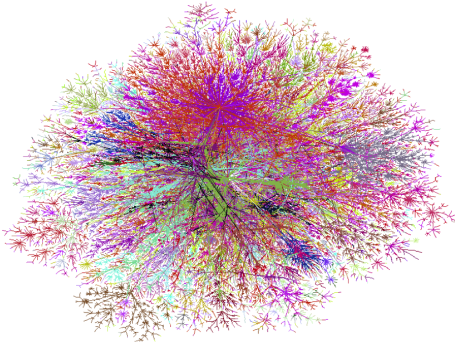
- friendship, collaborations, e-mail, etc.



Information networks

Nodes store information, links associate information

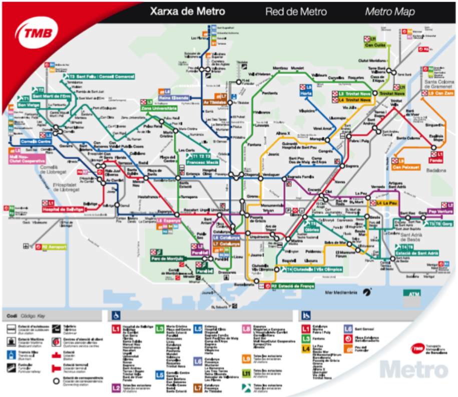
- citation networks, the web, p2p networks, etc.



Technological networks

Man-built for the distribution of a commodity

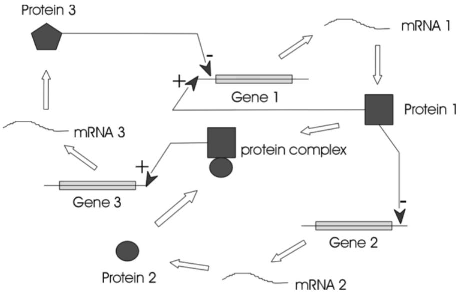
- telephone networks, power grids, transportation networks, etc.



Biological networks

Represent biological systems

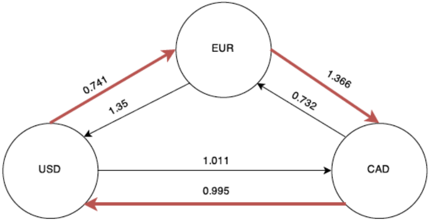
- protein-protein interaction networks, gene regulation networks, metabolic pathways, etc.



Financial networks

Nodes = financial assets, links = associated value or information

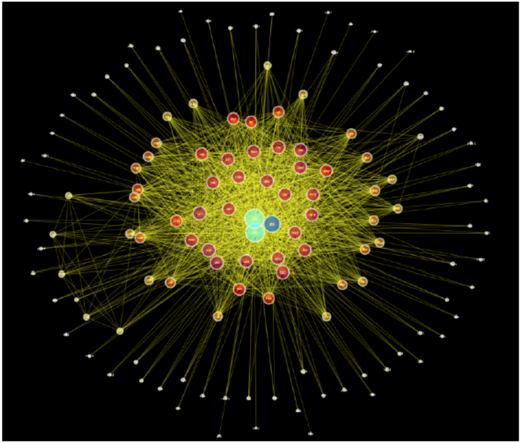
- Forex network I: Nodes = currencies, links = exchange value



- Forex network II: Nodes = currencies, links = nominal dollar value of all transactions between those two currencies (volume of trading)
  see: `http://ipeatunc.blogspot.com.es/2011/06/international-forex-network-1998-2010.html`

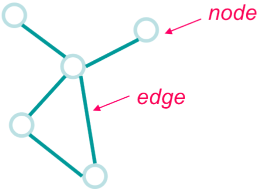
Financial networks

The Forex network (2015): Nodes = currencies, links = exchange value



Representing networks

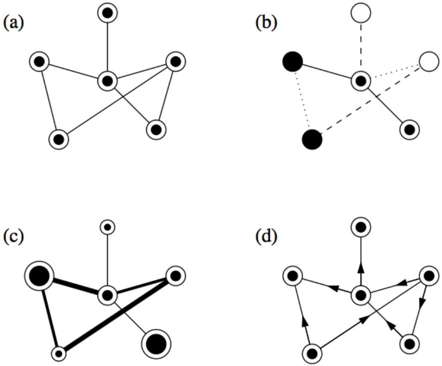
- Network $\equiv$ Graph
- Networks are just collections of "points" joined by "lines"



points	lines	
vertices	edges, arcs	math
nodes	links	computer science
sites	bonds	physics
actors	ties, relations	sociology

Types of networks

From [Newman, 2003]



(a) unweighted, undirected
(b) discrete vertex and edge types, undirected
(c) varying vertex and edge weights, undirected
(d) directed

Descriptive measures of networks

- real networks exhibit small diameter
- real networks have high clustering coefficient (or transitivity)
- real networks' degree distribution follows a power-law (i.e. are *scale free*)

From [Newman, 2003]

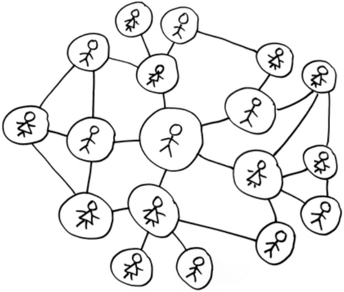
	network	type	n	m	z	ℓ	α	$C^{(1)}$	$C^{(2)}$	r	Ref(s).
social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208	20, 416
	company directors	undirected	7 673	55 392	14.44	4.60	–	0.59	0.88	0.276	105, 323
	math coauthorship	undirected	253 339	496 489	3.92	7.57	–	0.15	0.34	0.120	107, 182
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363	311, 313
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127	311, 313
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1				8, 9
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0		0.16		136
	email address books	directed	16 881	57 029	3.38	5.22	–	0.17	0.13	0.092	321
	student relationships	undirected	573	477	1.66	16.01	–	0.005	0.001	−0.029	45
	sexual contacts	undirected	2 810				3.2				265, 266
information	WWW nd.edu	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	−0.067	14, 34
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7				74
	citation network	directed	783 339	6 716 198	8.57		3.0/–				351
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	–	0.13	0.15	0.157	244
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44		119, 157
technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.035	0.39	−0.189	86, 148
	power grid	undirected	4 941	6 594	2.67	18.99	–	0.10	0.080	−0.003	416
	train routes	undirected	587	19 603	66.79	2.16	–		0.69	−0.033	366
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	−0.016	318
	software classes	directed	1 377	2 213	1.61	1.51	–	0.033	0.012	−0.119	395
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0	0.010	0.030	−0.154	155
	peer-to-peer network	undirected	880	1 296	1.47	4.28	2.1	0.012	0.011	−0.366	6, 354
biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	−0.240	214
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	−0.156	212
	marine food web	directed	135	598	4.43	2.05	–	0.16	0.23	−0.263	204
	freshwater food web	directed	92	997	10.84	1.90	–	0.20	0.087	−0.326	272
	neural network	directed	307	2 359	7.68	3.97	–	0.18	0.28	−0.226	416, 421

z mean deg; l mean distance; α exponent of deg. distrib. if power law; $C^{\cdot}$ clustering coef.

Small-world phenomenon

Low diameter and high transitivity

- Only 6 hops separate any two people in the world
- A friend of a friend is also frequently a friend



Measuring the small-world phenomenon, I

- Let d_{ij} be the shortest-path distance between nodes i and j
- To check whether "any two nodes are within 6 hops", we use:
  - The diameter (longest shortest-path distance) as

$$d = \max_{i,j} d_{ij}$$

  - The average shortest-path length as

$$l = \frac{2}{n\,(n-1)} \sum_{i>j} d_{ij}$$

  - The harmonic mean shortest-path length as

$$l^{-1} = \frac{2}{n\,(n-1)} \sum_{i>j} d_{ij}^{-1}$$

But..

- Can we mimic this phenomenon in simulated networks ("models")?
- The answer is YES!

The (basic) random graph model

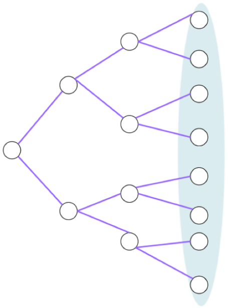
a.k.a. ER model

Basic $G_{n,p}$ Erdös-Rényi random graph model:

- parameter n is the number of vertices
- parameter p is s.t. $0 \leq p \leq 1$
- Generate and edge (i,j) independently at random with probability p

Measuring the diameter in ER networks

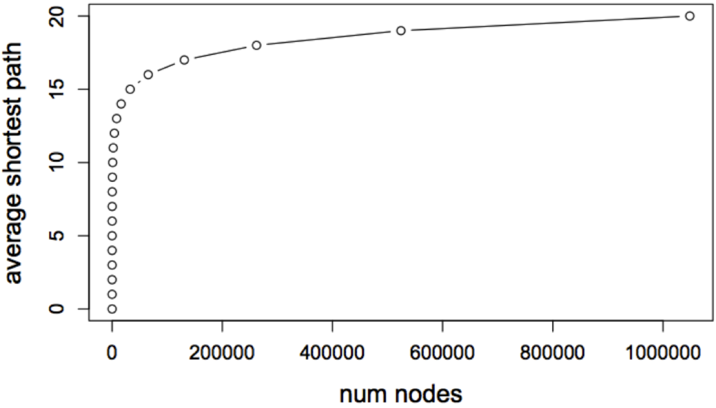
Want to show that the diameter in ER networks is small



- Let the average degree be z
- At distance l, can reach z^l nodes
- At distance $\frac{\log n}{\log z}$, reach all n nodes
- So, diameter is (roughly) $O(1)$
  (Show that $z = (n-1)p$)

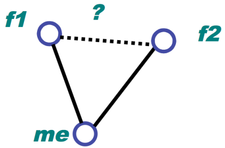
ER networks have small diameter

As shown by the following simulation



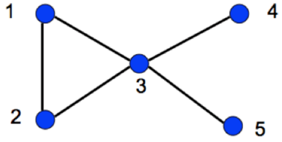
Measuring the small-world phenomenon, II

- To check whether "the friend of a friend is also frequently a friend", we use:
  - The transitivity or clustering coefficient, which basically measures the probability that two of my friends are also friends



Global clustering coefficient

$$C = \frac{3 \times \text{number of triangles}}{\text{number of connected triples}}$$



$$C = \frac{3 \times 1}{8} = 0.375$$

Local clustering coefficient

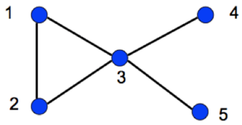
- For each vertex i, let n_i be the number of neighbors of i
- Let C_i be the fraction of pairs of neighbors that are connected within each other

$$C_i = \frac{\text{nr. of connections between } i\text{'s neighbors}}{\frac{1}{2} n_i \, (n_i - 1)}$$

- Finally, average C_i over all nodes i in the network

$$C = \frac{1}{n} \sum_i C_i$$

Local clustering coefficient example



- $C_1 = C_2 = 1/1$
- $C_3 = 1/6$
- $C_4 = C_5 = 0$
- $C = \frac{1}{5}(1 + 1 + 1/6) = 13/30 = 0.433$

ER networks do not show transitivity

- $C = p$, since edges are added independently
- Given a graph with n nodes and m edges, we can "estimate" p as

$$\hat{p} = \frac{m}{1/2 \; n \; (n-1)}$$

- We say that clustering is high if $C \gg \hat{p}$
  - Hence, ER networks do not have high clustering coefficient since for them $C \approx \hat{p}$

ER networks do not show transitivity

Table 1: Clustering coefficients, C, for a number of different networks; n is the number of node, z is the mean degree. Taken from [146].

Network	n	z	C measured	C for random graph
Internet [153]	6,374	3.8	0.24	0.00060
World Wide Web (sites) [2]	153,127	35.2	0.11	0.00023
power grid [192]	4,941	2.7	0.080	0.00054
biology collaborations [140]	1,520,251	15.5	0.081	0.000010
mathematics collaborations [141]	253,339	3.9	0.15	0.000015
film actor collaborations [149]	449,913	113.4	0.20	0.00025
company directors [149]	7,673	14.4	0.59	0.0019
word co-occurrence [90]	460,902	70.1	0.44	0.00015
neural network [192]	282	14.0	0.28	0.049
metabolic network [69]	315	28.3	0.59	0.090
food web [138]	134	8.7	0.22	0.065

So ER networks do not have high clustering, but..

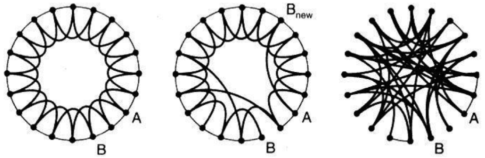
- Can we mimic this phenomenon in simulated networks ("models"), while keeping the diameter small?
- The answer is YES!

The Watts-Strogatz model, I

From [Watts and Strogatz, 1998]

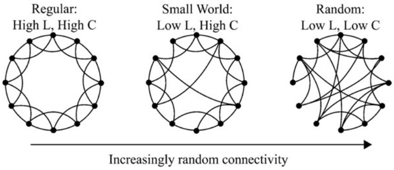
Reconciling two observations from real networks:

- High clustering: my friend's friends are also my friends
- small diameter



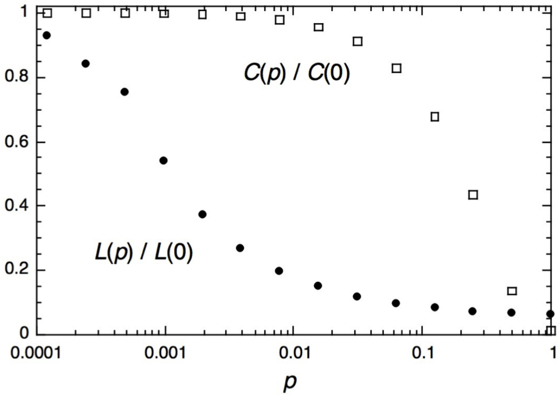
The Watts-Strogatz model, II

- Start with all n vertices arranged on a ring
- Each vertex has intially 4 connections to their closest nodes
  - mimics local or geographical connectivity
- With probability p, rewire each local connection to a random vertex
  - $p = 0$ high clustering, high diameter
  - $p = 1$ low clustering, low diameter (ER model)
- What happens in between?
  - As we increase p from 0 to 1
    - Fast decrease of mean distance
    - Slow decrease in clustering



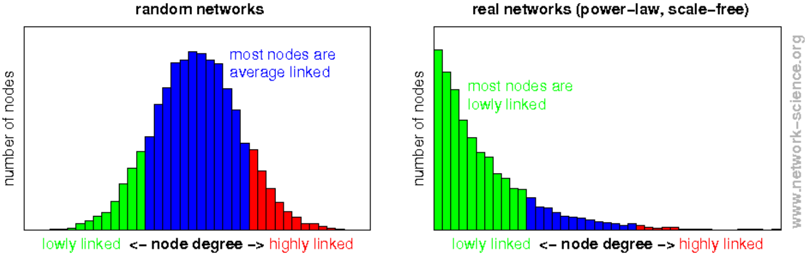
The Watts-Strogatz model, III

For an appropriate value of $p \approx 0.01$ (1%), we observe that the model achieves high clustering and small diameter



Degree distribution

Histogram of nr of nodes having a particular degree

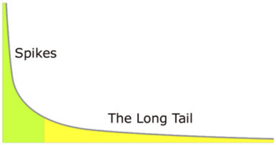


$$f_k = \text{fraction of nodes of degree } k$$

Scale-free networks

The degree distribution of most real-world networks follows a power-law distribution

$$f_k = ck^{-\alpha}$$



- "heavy-tail" distribution, implies existence of hubs
- hubs are nodes with very high degree

Random networks are not scale-free!

For random networks, the degree distribution follows the binomial distribution (or Poisson if n is large)

$$f_k = \binom{n}{k} p^k (1-p)^{(n-k)} \approx \frac{z^k e^{-z}}{k!}$$

- Where $z = p(n-1)$ is the mean degree
- Probability of nodes with very large degree becomes exponentially small
  - so no hubs

So ER networks are not scale-free, but..

- Can we obtained scale-free simulated networks?
- The answer is YES!

Preferential attachment

- "Rich get richer" dynamics
  - The more someone has, the more she is likely to have
- Examples
  - the more friends you have, the easier it is to make new ones
  - the more business a firm has, the easier it is to win more
  - the more people there are at a restaurant, the more who want to go

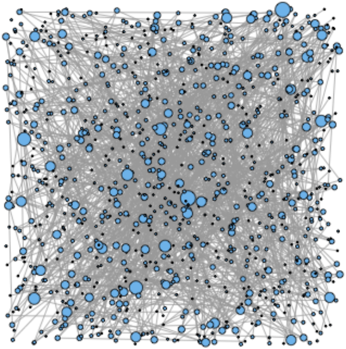
Barabási-Albert model

From [Barabási and Albert, 1999]

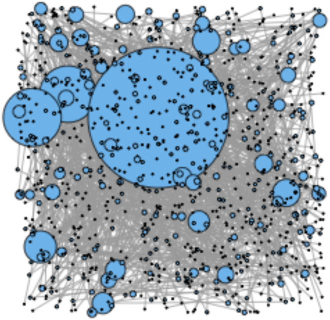
- "Growth" model
  - The model controls how a network grows over time
- Uses preferential attachment as a guide to grow the network
  - new nodes prefer to attach to well-connected nodes
- (Simplified) process:
  - the process starts with some initial subgraph
  - each new node comes in with m_0 edges
  - probability of connecting to existing node i is proportional to i's degree
  - results in a power-law degree distribution with exponent $\alpha = 3$

ER vs. BA

Experiment with 1000 nodes, 999 edges ($m_0 = 1$ in BA model).



random



preferential attachment

In summary..

phenomenon	real networks	ER	WS	BA
small diameter	yes	yes	yes	yes
high clustering	yes	no	yes	yes[1]
scale-free	yes	no	no	yes

[1] clustering coefficient is higher than in random networks, but not as high as for example in WS networks

References I

Barabási, A.-L. and Albert, R. (1999).
Emergence of scaling in random networks.
science, 286(5439):509–512.

Baronchelli, A., i Cancho, R. F., Pastor-Satorras, R., Chater, N., and Christiansen, M. H. (2013).
Networks in cognitive science.
Trends in cognitive sciences, 17(7):348–360.

Barrat, A., Barthelemy, M., and Vespignani, A. (2008).
Dynamical processes on complex networks, volume 1.
Cambridge University Press Cambridge.

References II

Kolaczyk, E. D. (2009).
Statistical analysis of network data.
Springer.

Newman, M. (2009).
Networks: an introduction.
Oxford University Press.

Newman, M. E. (2003).
The structure and function of complex networks.
SIAM review, 45(2):167–256.

Watts, D. J. and Strogatz, S. H. (1998).
Collective dynamics of small-worldnetworks.
nature, 393(6684):440–442.