Social Networks

Pierre Senellart

Collège de France, 30 May 2012

Most popular Web sites

1	google.com
2	facebook.com
3	youtube.com
4	yahoo.com
5	baidu.com
6	wikipedia.org
7	live.com
8	twitter.com
9	qq.com
10	amazon.com
11	blogspot.com
12	linkedin.com
13	google.co.in
14	taobao.com
15	sina.com.cn
16	yahoo.co.jp
17	msn.com
18	wordpress.com
19	google.com.hk
20	t.co
21	google.de
22	ebay.com
23	google.co.jp
24	googleusercontent.com
25	google.co.uk
26	yandex.ru
27	163.com
28	weibo.com

(Alexa)

Most popular Web sites

1. google.com
2. facebook.com
3. youtube.com
4. yahoo.com
5. baidu.com
6. wikipedia.org
7. live.com
8. twitter.com
9. qq.com
10. amazon.com
11. blogspot.com
12. linkedin.com
13. google.co.in
14. taobao.com
15. sina.com.cn
16. yahoo.co.jp
17. msn.com
18. wordpress.com
19. google.com.hk
20. t.co
21. google.de
22. ebay.com
23. google.co.jp
24. googleusercontent.com
25. google.co.uk
26. yandex.ru
27. 163.com
28. weibo.com

(Alexa)

Social networking sites

Most popular Web sites

1	google.com
2	facebook.com
3	youtube.com
4	yahoo.com
5	baidu.com
6	wikipedia.org
7	live.com
8	twitter.com
9	qq.com
10	amazon.com
11	blogspot.com
12	linkedin.com
13	google.co.in
14	taobao.com
15	sina.com.cn
16	yahoo.co.jp
17	msn.com
18	wordpress.com
19	google.com.hk
20	t.co
21	google.de
22	ebay.com
23	google.co.jp
24	googleusercontent.com
25	google.co.uk
26	yandex.ru
27	163.com
28	weibo.com

(Alexa)

Social networking sites

Sites with social networking features (friends, user-shared content, user profiles, etc.)

Social data on the Web

Huge numbers of users:

Facebook	900 million
QQ	540 million
W. Live	330 million
Weibo	310 million
Google+	170 million
Twitter	140 million
LinkedIn	100 million

Social data on the Web

Huge numbers of users:

Facebook	900 million
QQ	540 million
W. Live	330 million
Weibo	310 million
Google+	170 million
Twitter	140 million
LinkedIn	100 million

Huge volume of shared data:

250 million tweets per day on Twitter (3,000 per second on average!)...

... including statements by heads of states, revelations of political activists, etc.

Exploiting social Web data

Very rich source of information, lots that can be done with it: technology watch, sentiment analysis, sociological analysis, etc.

Many challenges as well: unbiased sampling from social networks? how to keep up to date with 3,000 tweets every second? how to manage the petabytes of data of social networking sites?

Focus on two problems:

- (Primarily) What is the structure of social networks? How to model them?
- Socially aware Web search

Outline

Small worlds

I proposed a more difficult problem: to find a chain of contacts linking myself with an anonymous riveter at the Ford Motor Company — and I accomplished it in four steps. The worker knows his foreman, who knows Mr. Ford himself, who, in turn, is on good terms with the director general of the Hearst publishing empire. I had a close friend, Mr. Árpád Pásztor, who had recently struck up an acquaintance with the director of Hearst Publishing. It would take but one word to my friend to send a cable to the general director of Hearst asking him to contact Ford who could in turn contact the foreman, who could then contact the riveter, who could then assemble a new automobile for me, would I need one.

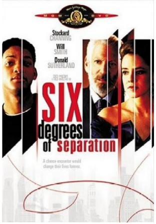
[...] Our friend was absolutely correct: nobody from the group needed more than five links in the chain to reach, just by using the method of acquaintance, any inhabitant of our Planet.

[Karinthy, 1929]

TELECOM ParisTech

Six degrees of separation

- Idea that two persons on Earth are separated by a chain of six individuals who know each other
- Appears widely in popular culture:



It's a small world!

TELECOM ParisTech

Stanley Milgram's experiment [Travers and Milgram, 1969]

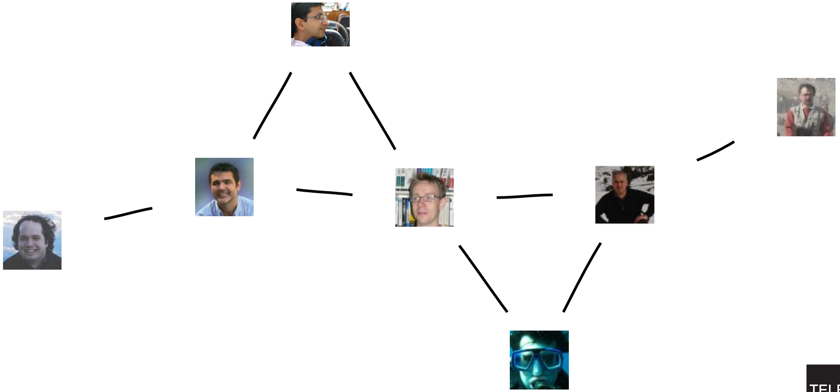
© Al Satterwhite

- Stanley Milgram (1933-1984): social psychologist
- Experiment: people are asked to send a message to some unknown person, by forwarding it to an acquaintance who might be closer to this person
- Results: only 29% of the messages arrived, with a mean number of acquaintances of 5.2.
- Validates somehow the 6-degree theory!
- Other more recent experiments [Dodds et al., 2003] confirm this order of magnitude.

Simple model of a social network

A social network is just a graph:

- individuals, data items, groups, etc., are nodes
- connections are (possibly directed and labeled) edges



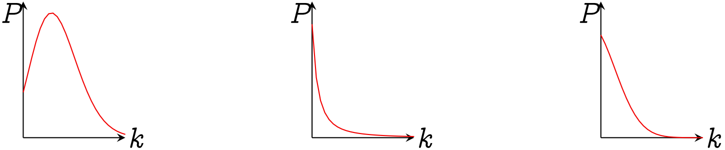
Characteristics of interest of a network

Sparsity. Is the network sparse ($|A| \ll |S|^2$)?

Typical distance. What is the mean distance between any pairs of vertices?

Local clustering. If a is connected to both b and c, is the probability that b is connected to c significantly greater than the probability any two nodes are connected?

Degree distribution. What is the distribution of the degree of vertices?



Poisson	Power-law	Gaussian
$\frac{\lambda^k}{k!}$	$k^{-\gamma}$	e^{-k^2}

Characteristics of social networks

- Choose any social network: real-life acquaintance network [Amaral et al., 2000], Twitter follower graph [Kwak et al., 2010], scientific collaboration network [Amaral et al., 2000], romantic or sexual relationships [Amaral et al., 2000, Liljeros et al., 2001], etc.

Characteristics of social networks

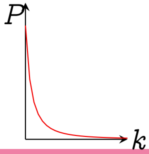
- Choose any social network: real-life acquaintance network [Amaral et al., 2000], Twitter follower graph [Kwak et al., 2010], scientific collaboration network [Amaral et al., 2000], romantic or sexual relationships [Amaral et al., 2000, Liljeros et al., 2001], etc.
- Sparse graph.

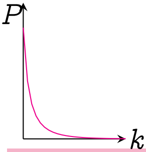
Characteristics of social networks

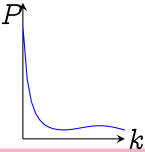
- Choose any social network: real-life acquaintance network [Amaral et al., 2000], Twitter follower graph [Kwak et al., 2010], scientific collaboration network [Amaral et al., 2000], romantic or sexual relationships [Amaral et al., 2000, Liljeros et al., 2001], etc.
- Sparse graph.
- Strong local clustering (well, not for romantic relationships)

Characteristics of social networks

- Choose any social network: real-life acquaintance network [Amaral et al., 2000], Twitter follower graph [Kwak et al., 2010], scientific collaboration network [Amaral et al., 2000], romantic or sexual relationships [Amaral et al., 2000, Liljeros et al., 2001], etc.
- Sparse graph.
- Strong local clustering (well, not for romantic relationships)
- Degree distribution: usually a power-law with γ between 2 and 3. Sometimes slightly modified: exponential or Gaussian cut-off (real-life acquaintances) or, on the contrary, more high-degree nodes than expected (Twitter). Sometimes even more complicated [Sala et al., 2010b].







Other kinds of networks

Similar characteristics for:

- Neural network of a worm [Watts and Strogatz, 1998]
- Metabolic interaction network [Jeong et al., 2000]
- The Internet [Faloutsos et al., 1999]
- The World Wide Web [Broder et al., 2000]

Counter-examples: networks with a 2D embedding, such as road networks.

Other kinds of networks

Similar characteristics for:

- Neural network of a worm [Watts and Strogatz, 1998]
- Metabolic interaction network [Jeong et al., 2000]
- The Internet [Faloutsos et al., 1999]
- The World Wide Web [Broder et al., 2000]

Counter-examples: networks with a 2D embedding, such as road networks.

Why?

Random networks [Solomonoff and Rapoport, 1951, Erdős and Rényi, 1960]

Construction

1. Start with n vertices and a probability p. Assume $p > \frac{1}{n}$.
2. For each pair of vertices (u, v), insert an edge between u and v with probability p.

Random networks [Solomonoff and Rapoport, 1951, Erdős and Rényi, 1960]

Network characteristics

- Sparse if $p \ll 1$

Constr

1. St
2. Fo ... v

   w

Random networks [Solomonoff and Rapoport, 1951, Erdős and Rényi, 1960]

Network characteristics

- Sparse if $p \ll 1$
- Logarithmic typical distance (inside the giant connected component)!

Constr

1. St
2. Fo ... v

w

Random networks [Solomonoff and Rapoport, 1951, Erdős and Rényi, 1960]

Network characteristics

- Sparse if $p \ll 1$
- Logarithmic typical distance (inside the giant connected component)!
- No local clustering.

Constr

1. St
2. Fo . v

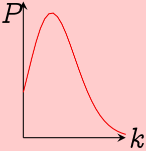
   w

TELECOM ParisTech

Random networks [Solomonoff and Rapoport, 1951, Erdős and Rényi, 1960]

Network characteristics

- Sparse if $p \ll 1$
- Logarithmic typical distance (inside the giant connected component)!
- No local clustering.
- Poisson degree distribution



Constr

1. St
2. Fo . v

w

Preferential attachment [Barabási and Albert, 1999]

Construction

1. Start with a small graph of size m_0, let m be a constant with $m < m_0$.
2. One after the other, $n - m_0$ vertices are added to the graph, connecting them to m existing vertices; the probability of connecting to a vertex is proportional to its degree.

Rich get richer!

Preferential attachment [Barabási and Albert, 1999]

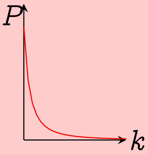
Constr

1. St
   m

2. O
   co
   co

Network characteristics

- Sparse if m and n are chosen appropriately.
- Small typical distance.
- Strong local clustering
- Power-law degree distribution (actually, with $\gamma = 3$, but variations allow arbitrary exponents).



So why not a full power-law distribution?

- Exponential or Gaussian people (real-life networks): in the real life, there is a cost to establish connections, cannot have too many of them. In addition, people die and stop making connections.
- More connections than expected (Twitter): celebrities have a special status on Twitter and attract more followers than the preferential attachment model predicts.

What have we learned?

- Nothing surprising about the small-world effect! Already happens in a completely random graph: due to random connections.
- Clustering, degree distribution: partly explained by the history of the network, and by the fact that rich get richer (preferential attachment)
- Preferential attachment is not the last word on this. More refined models do exist! [Sala et al., 2010a, Vazquez, 2003]
- We know how to build reasonable synthetic social networks.

Outline

Social Networking Sites

Social Networks: Structure and Models

Socially Aware Web Search

Conclusion

Why social search?

- Many Web queries are ambiguous or too broad to retrieve what the user is looking for: "music", "president", "Paris"
- General idea: my friends' interest are correlated to mine.
- If I bias my search results based on my social network:
  - I can be less precise in my queries; when I talk about spouse it is by default my spouse, company my company, etc.
  - I can get info about my friends
  - I can get sentiment/opinion/taste (e.g., about a new gadget) biased by those of of people like me (even if I don't know them)
- Risks:
  - Remove objectivity in Web search – what I want to hear and not the truth
  - Community withdrawal

Information retrieval with social scoring [Schenkel et al., 2008]

- Setting: multi-partite graph, e.g., Flickr (user-photo-tag)
- Social weighting:
  - Given a friendship relation $F(u, u')$ (explicit or implicit) between two users, computer an extended friendship relation:

$$\tilde{F}(u, u') = \frac{\alpha}{|U|} + (1-\alpha) \max_{\text{chemin } u = u_0 \dots u_k = u'} \prod_{i=0}^{k-1} F(u_i, u_{i+1})$$

    ($0 < \alpha < 1$ constant; $|U|$: number of users)
  - Instead of using a global weighting $\text{tf-idf}(t, d) = \text{tf}(t, d) \times \text{idf}(t, d)$ use a social weighting dependent of u:

$$\text{tf-idf}_u(t, d) = \left(\sum_{u' \in U} F(u, u') \cdot \text{tf}_{u'}(t, d) \right) \times \text{idf}(t, d)$$

Top-k with social score [Benedikt et al., 2008]

- Possible to use refined algorithms to get only top-k best results (as in classical Web search)...
- ... but this requires precomputing tf-idf$_u(t, d)$ for each user: impossible
- To avoid this:
  1. Partition the graph of users in clusters of very similar users
  2. Use the scores inside each cluster as estimations of whether the top-k results found are the best
  3. $\Rightarrow$ gives approximate results, but good enough

Outline

TELECOM ParisTech

Many challenges, many applications

- better models of social graphs (taking into account locality, dynamics, etc.)
- applications of network models: epidemiology simulations, propagation of rumors, resilience to censorship
- mass of social data on the Web, waiting to be exploited:
  - better search results, better recommendation, better understanding of our world
  - monitoring the Web for things that may interest me
  - intrusive advertising, extreme profiling (determine whether you're gay, pregnant, or activist by looking at your social network or social data)
- a wide collection of social algorithms
- need for data sharing models in social networks, taking into account privacy, distribution, etc. – Webdamlog

To go further

[Watts, 2003]: an easy-to-read book describing the area of network science, including models and concrete applications

[Newman et al., 2006]: an in-detail survey of the most fundamental works on network theory, networks models, and experimentations on real-world networks

Thanks to Serge Abiteboul, Silviu Maniu, Yann Ollivier.

Bibliography I

L. A. Amaral, A. Scala, M. Barthelemy, and H. E. Stanley. Classes of small-world networks. *PNAS*, 97(21):11149–11152, October 2000.

Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, October 1999.

Michael Benedikt, Sihem Amer Yahia, Laks Lakshmanan, and Julia Stoyanovich. Efficient network-aware search in collaborative tagging sites. In *Proc. VLDB*, Auckland, New Zealand, August 2008.

Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, and Janet Wiener. Graph structure in the web. *Computer Networks*, 33(1-6): 309–320, 2000.

Bibliography II

Peter Sheridan Dodds, Roby Muhamad, and Duncan J. Watts. An experimental study of search in global social networks. *Science*, 301 (5634):827–829, August 2003.

P. Erdős and A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5:17–61, 1960.

M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. In *Proc. SIGCOMM*, Cambridge, USA, August 1999.

H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabasi. The large-scale organization of metabolic networks. *Nature*, 407(6804), 2000.

Bibliography III

Frigyes Karinthy. Chains. In *Everything is different*. 1929. Translated from Hungarian by Ádám Makkai, as reproduced in [Newman et al., 2006].

Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue B. Moon. What is Twitter, a social network or a news media? In *Proc. WWW*, Raleigh, NC, USA, 2010.

F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Aaberg. The web of human sexual contacts. *Nature*, 411(6840): 907–908, 2001.

Mark Newman, Albert-László Barabási, and Duncan J. Watts. *The Structure and Dynamics of Networks*. Princeton University Press, 2006.

Bibliography IV

Alessandra Sala, Lili Cao, Christo Wilson, Robert Zablit, Haitao Zheng, and Ben Y. Zhao. Measurement-calibrated graph models for social network experiments. In *Proc. WWW*, Raleigh, NC, USA, 2010a.

Alessandra Sala, Haitao Zheng, Ben Y. Zhao, Sabrina Gaito, and Gian Paolo Rossi. Brief announcement: revisiting the power-law degree distribution for social graph analysis. In *PODC*, Zurich, Switzerland, 2010b.

Ralf Schenkel, Tom Crecelius, Mouna Kacimi, Sebastian Michel, Thomas Neumann, Josiane X. Parreira, and Gerhard Weikum. Efficient top-k querying over social-tagging networks. In *Proc. SIGIR*, Singapore, Singapore, July 2008.

Bibliography V

Ray Solomonoff and Anatol Rapoport. Connectivity of random nets. *Bulletin of Mathematical Biology*, 13(2):107–117, June 1951.

Jeffrey Travers and Stanley Milgram. An experimental study of the small world problem. *Sociometry*, 34(4), December 1969.

A. Vazquez. Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Physical Review E*, 67(056104), 2003.

Duncan J. Watts. *Six Degrees: The Science of a Connected Age*. W. V. Norton & Company, 2003.

Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.