Network Science: From Social Networks to Human Mobility

Albert-László Barabási

Center for Complex Networks Research
Northeastern University
Department of Medicine and CCSB
Harvard Medical School

www.BarabasiLab.com
natural phenomena
controlled, predicted, quantified, understood, described,
quantified, understood, described, quantified, predicted, controlled.
Tropical forest in Yucatan, Mexico

“Lévy flight” by a spider monkey
We need images to illustrate these:

- We can predict where an electron will go, but we cannot foresee and stop economic crises.
- We can turn a gene on or off, but we have no control over wars and battles.
- We can send a robot to Mars, but we are lost if asked to predict phenomena we would think we should know the most about, which is the actions of our fellow humans.
We need images to illustrate these:

- We can predict where an electron will go, but we cannot foresee and stop economic crises.
- We can turn a gene on or off in... but we have no control over wars and battles.
- We can send a robot to Mars, but we are lost if asked to predict phenomena we would think we should know the most about, which is the actions of our fellow humans.
to predict anything, you need data. Lots of data.
Why would a physicist care about human behavior?
Erdös-Rényi model (1960)

Connect with probability $p$

$p = 1/6$

$N = 10$

$\langle k \rangle \sim 1.5$

Pál Erdös (1913-1996)

- Democratic
- Random

Poisson distribution
Erdös–Rényi model (1960)

Democratic Random

Connect with probability $p = \frac{1}{6}$

$N = 10$

$\langle k \rangle \sim 1.5$

Poisson distribution
World Wide Web

**Nodes**: WWW documents

**Links**: URL links

Over 10 billion documents

**ROBOT**: collects all URL’s found in a document and follows them recursively

\[ P(k) \sim k^{-\gamma} \]

World Wide Web

Over 10 billion documents

ROBOT:
collects all URL's found in a document and follows them recursively

Nodes: WWW documents

Links: URL links


Exponential Network

Scale-free Network

\[ P(k) \sim k^{-\gamma} \]
INTERNET BACKBONE

**Nodes**: computers, routers

**Links**: physical lines

(Faloutsos, Faloutsos and Faloutsos, 1999)
Nodes: scientist (authors)
Links: write paper together

(SCIENCE COAUTHORSHIP

(Newman, 2000, Barabási et al 2001)
**Nodes**: scientist (authors)

**Links**: write paper together

(Newman, 2000, Barabási et al 2001)
Origin of SF networks: Growth and preferential attachment

(1) Networks continuously expand by the addition of new nodes
   - **WWW**: addition of new documents

(2) New nodes prefer to link to highly connected nodes.
   - **WWW**: linking to well known sites

**GROWTH:**
add a new node with m links

**PREFERENTIAL ATTACHMENT:** the probability that a node connects to a node with \( k \) links is proportional to \( k \).

\[
\Pi(k_i) = \frac{k_i}{\sum_j k_j}
\]
Origin of SF networks: Growth and preferential attachment

(1) Networks continuously expand by the addition of new nodes
   \[ \text{GROWTH: add a new node with } m \text{ links} \]
   \[ \text{WWW : addition of new documents} \]

(2) New nodes prefer to link to highly connected nodes.
   \[ \text{PREFERENTIAL ATTACHMENT: the probability that a node connects to a node with } k \text{ links is proportional to } k. \]
   \[ \Pi(k_i) = \frac{k_i}{\sum_i k_i} \]

Metabolic Network  
Protein Interactions

Robustness

Complex systems maintain their basic functions even under errors and failures (cell $\rightarrow$ mutations; Internet $\rightarrow$ router breakdowns)

\[ S \]

Fraction of removed nodes, $f$

$S$ vs. $f$

node failure
Robustness of scale-free networks

\[ \gamma \leq 3 : f_c = 1 \]

(R. Cohen et al PRL, 2000)

CONNECTING KNOWLEDGE
Understanding human trajectories

Center of Mass:
\[ \vec{r}_{cm} = \frac{1}{n_p} \sum_{i=1}^{n_p} \vec{r}_i. \]

Radius of Gyration:
\[ r_g^a(t) = \sqrt{\frac{1}{n_c^a(t)} \sum_{i=1}^{n_c^a} (\vec{r}_i^a - \vec{r}_{cm}^a)^2} \]
Scaling in human trajectories

\[ \beta_r = 1.65 \pm 0.15 \]
Most people travel very little on a daily basis—somewhere between 2-5 kms.

A few individuals regularly cover hundreds of kms.
Mobility Networks
Mobility Networks

$N = 22$

$N = 76$
Entropy Distribution Across the Population

$S_{\text{max}} \sim 6 \rightarrow$ a random user could be found in any of $2^{S_{\text{max}}} \sim 64$ locations.

Song, Qu, Blumm, Barabasi, Science (2010)
$S_{\text{max}} \approx 6 \rightarrow$ a random user could be found in any of $2^{S_{\text{max}}} \approx 64$ locations.

$S = 0.8 \rightarrow$ the real uncertainty in the user’s whereabouts is $2^{0.8} = 1.74$.

Song, Qu, Blumm, Barabasi, *Science* (2010)
Predictability $\Pi$ = success rate of the best predictive algorithm

Fano inequality relates entropy & predictability

$$S \leq -\Pi \log_2 \Pi - (1 - \Pi) \log_2 (1 - \Pi) + (1 - \Pi) \log_2 (N - 1)$$

$$\Pi \leq \Pi^{\text{max}} (S, N)$$

$\Pi^{\text{max}}$ = maximal predictability
Predictability $\Pi = \text{success rate of the best predictive algorithm}$

Fano inequality relates entropy & predictability

$S \leq -\Pi \log_2 \Pi - (1 - \Pi) \log_2 (1 - \Pi) + (1 - \Pi) \log_2 (N - 1)$

$\Pi \leq \Pi_{\text{max}} (S, N)$

$\Pi_{\text{max}} = \text{maximal predictability}$
93% predictability in our future whereabouts.

Hardly anyone under 80%

We are all equally predictable.

Song, Qu, Blumm, Barabasi, Science (2010)
Song, Qu, Blumm, Barabasi, *Science* (2010)
Origin of the high predictability

High spatial and temporal regularity:

- Divide user’s location time series into $24 \times 7 = 168$ hours/time frame.
- Define regularity $R$ as the probability of finding the user in his most visited location at a given time frame.

$R \sim 0.9$ at night!

$R \sim 0.7$ on average

$R \leq \prod$ Regularity is a generous lower bound of predictability.
With whom?
Most hubs. Or others similar to us.

Where?
Mostly nearby. As we rarely go too far.

Why?
Because we did it yesterday.