# Social networks and homophily

## 10th International Winter School on Inequality and Social Welfare Theory

Paolo Pin

Università di Siena

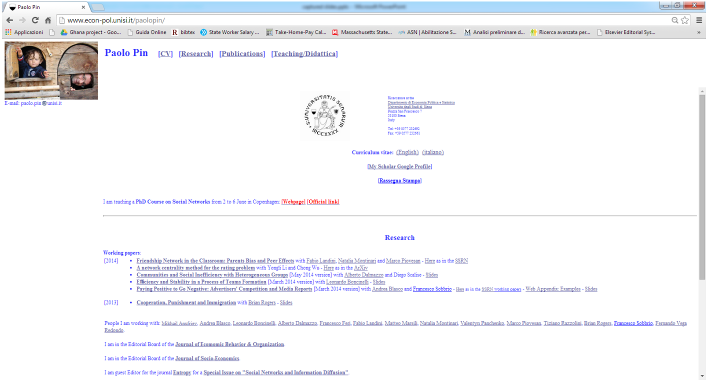
Canazei – January, 2015

UNIVERSITÀ DI SIENA 1240

# My coordinates

http://www.econ-pol.unisi.it/paolopin/

Mail: paolo.pin@unisi.it




UNIVERSITÀ DI SIENA 1240

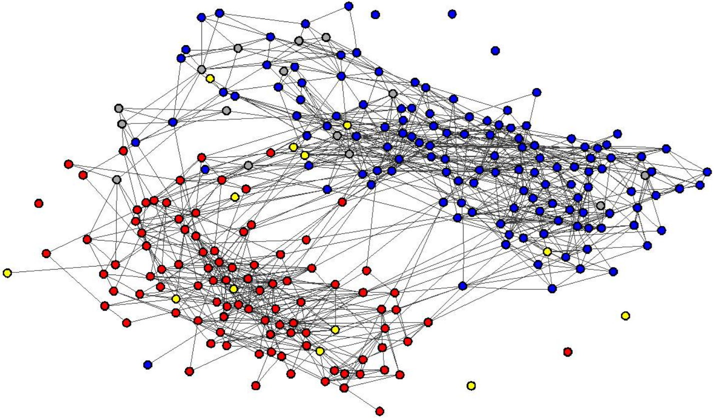
# Today's lecture



UNIVERSITÀ DI SIENA 1240

# An example: Friendships

Nodes are students from a US High School, there is a link if in a survey one cites the other as his/her friend



# AddHealth data

The National Longitudinal Study of Adolescent Health (Add Health)

- 1994 survey in 84 American high schools and middle schools, by the UNC Carolina Population Center
- They could nominate their friends from a list of all the other students in their school
- We consider a link whenever at least one of the two students nominate the other one
- Students were self–reporting a lot of other information in a huge survey

# Standard economic assumptions

- Completeness of markets: anyone can trade with anyone else
  - but most markets are not centralized
  - there are communication costs and asymmetric information risks when changing partners
  - this is not only the case in undeveloped markets
  - but it is also not the case in international economics
- Efficiency of markets: prices summarize all the information
  - the 2007/2009 crises is a good example of how limited communication flows moved prices away from the fundamentals
- *Homo economicus*
  - in taking decisions, economic agents are not influenced by peers
- Externalities: public goods and bads are global concepts
  - most of the externalities are instead local: because of geographic, but also cultural, distances

# Only frictions?

We observe networks in trading, information passing...

- But is a just by chance?
- Is the cost of changing partners just a negligible friction?
- Can we exclude any path dependence from the links that have been in place in the past?

If the answer to those questions is always ‘yes’, then networks are not important for economic theory

If some answer are ‘no’, then it is important to study how this constraints work, and what are their implications

# A classical historical example: centrality

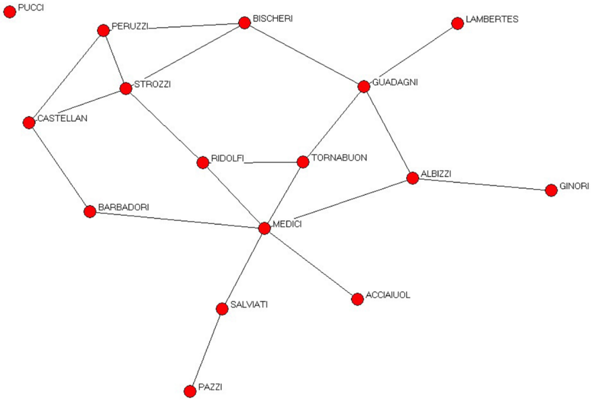



Figure 1: 15th Century Florentine Marriges (Padgett and Ansell (1993))

UNIVERSITÀ DI SIENA 1240

# Nodes and links

- $N = \{1, 2, \ldots, n\}$ nodes, vertices, agents, actors, players
- edges, links, ties: connections between nodes
  - They may have intensity (weighted)
    - How many hours do two people spend together per week?
    - How much of one country's GDP is traded with another?
  - They may just be 0 or 1 (unweighted)
    - Have two researchers written an article together?
    - Are two people "friends" on some social platform?
  - They may be "undirected" or "directed"
    - coauthors, friends,Facebook friends. . . are mutual relationships
    - link from on web page to another, citations, following on Twitter. . . one way

*Network architecture*: what is invariant under permutation of nodes

UNIVERSITÀ DI SIENA 1240

# Representation of links

Any characteristic can be associated to nodes

Links (as $i \to j$) can be represented

- as a list of couples of nodes: $g \subseteq N \times N$, with $ij \in g$ iff $i \to j$
- as an *adjacency matrix*: $g \in \{0,1\}^{n \times n}$ where $g_{ij} = 1$ iff $i \to j$

How would the weighted case be?

A network is a couple: $(N, g)$

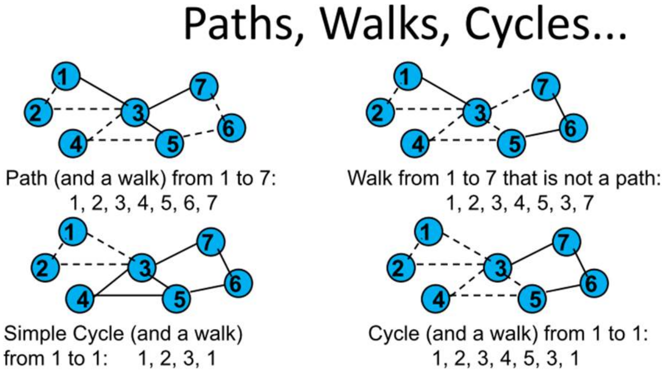
**Theory:** the two definitions are equivalent, the adjacency matrix is more compact

**Practice:** Facebook has $\sim 3.10^9$ users and $\sim 10^{11}$ links...
an adjacency matrix requires $\sim 10^{19}$ entries!

UNIVERSITÀ DI SIENA 1240

# Walks on networks

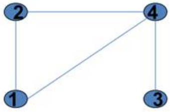
Other notions: *walk*, *path*, *cycle*, *tree*, (geodesic) *distance*, *infinite distance*




UNIVERSITÀ DI SIENA 1240

# Counting walks

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




$$g^2=\begin{pmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 3 \end{pmatrix}$$

number of walks of length 2 from i to j

$$g^3=\begin{pmatrix} 2 & 3 & 1 & 4 \\ 3 & 2 & 1 & 4 \\ 1 & 1 & 0 & 3 \\ 4 & 4 & 3 & 2 \end{pmatrix}$$

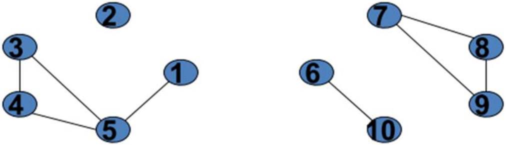
number of walks of length 3 from i to j

ITÀ DI SIENA 1240

# Components

- A network is *connected* if there is a path between every two nodes
- *Component*: maximal connected subgraph $(N', g')$ with $N' \subseteq N$

Example: a network with 4 components




UNIVERSITÀ DI SIENA 1240

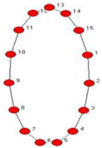
# Diameter

The *diameter* can be two different things:

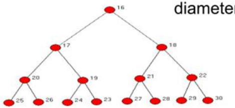
- The largest (geodesic) distance (i.e. largest shortest path)
  - If the network is unconnected, we usually take the largest finite distance
  - or the largest distance of the largest component
- The average (finite) path length

Examples:

K levels has $n = 2^{K+1}-1$ nodes
so, $K = \log_2(n+1) - 1$
diameter is 2K



diameter is either
n/2 or (n-1)/2



diameter is on order of
$2 \log_2(n+1)$

UNIVERSITÀ DI SIENA 1240

# Diameter for contagion processes

Consider a situation in which a node may be infected, and pass the infection to neighbors

- An extensively studied model in epidemiology: SIS
- Nodes are infected or susceptible
- Allows nodes to change behaviors back and forth over time (alternatives is SIR)
- Model of catching some recurring diseases, who to vote for, acquisition of information, viral effects etc.

- In this class of models the *overall* subsceptibility of the network depends only on its diameter (and on the rate of *infectiveness*)
- Intuition: how many steps does it take to the infection, if successfull, to reach any node of the network?

# Viral effects – the dynamics

The ‘how many steps’ question is contingent on: ‘if successfull’
To pass from 0 to full infection state may be a one–in–a–milion event

http://www.facegroup.com/how-stuff-spreads-1-gangnam-style-vs-harlem-shake.html

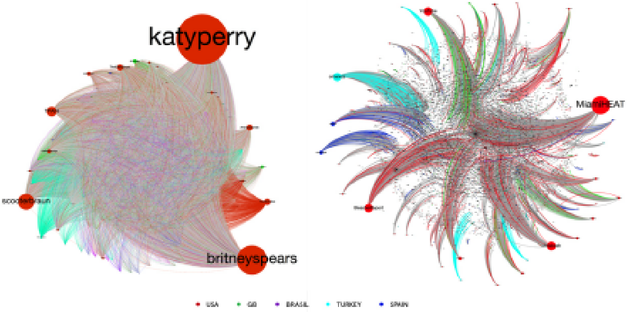



Figure 1: Diffusion of the Gangnam Style and the Harlem Shake videos on the internet (D’Orazio and Owens, 2013).

UNIVERSITÀ DI SIENA 1240

# Neighborhood

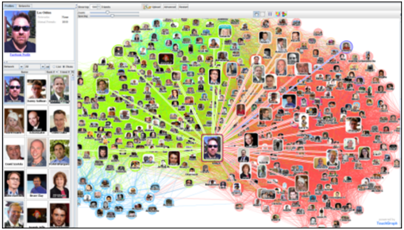
Neighborhood, friends, peers. . . are represented as a *neighborhood*

- $N_i(g) = \{j | ij \in g\}$
- or $N_i(g) = \{j | g_{ij} = 1\}$

- the last one allows for *second neighborhood*: $N_i(g) = \{j | g^2_{ij} \geq 1\}$, and so on

  Usually $i$ is never counted

Examples:



# Degree and degree distributions

The degree of a node is her/his number of friends: $d_i = |N_i(g)|$

From this we can plot degree distributions

Example from Barabasi & Albert (1999):

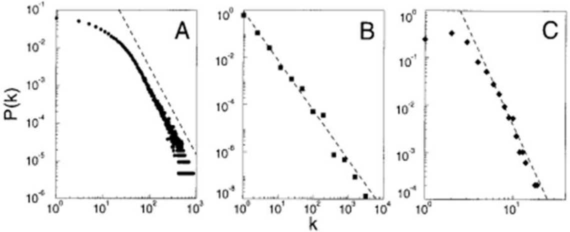



**Fig. 1.** The distribution function of connectivities for various large networks. **(A)** Actor collaboration graph with $N = 212{,}250$ vertices and average connectivity $\langle k \rangle = 28.78$. **(B)** WWW, $N = 325{,}729$, $\langle k \rangle = 5.46$ (6). **(C)** Power grid data, $N = 4941$, $\langle k \rangle = 2.67$. The dashed lines have slopes (A) $\gamma_{actor} = 2.3$, (B) $\gamma_{www} = 2.1$ and (C) $\gamma_{power} = 4$.

UNIVERSITÀ DI SIENA 1240

# Random graphs

# Example: a survey in a school

## Questionario

Compilate per favore il questionario, usando come riferimento la lista numerata di tutti gli studenti della scuola che avete in allegato. I dati serviranno a costruire le caratteristiche del network della vostra scuola (come quello in figura). Vi ricordiamo che la lista non sarà visibile a chi successivamente analizzerà i dati, e non sarete pertanto riconoscibili nemmeno guardando il network.

INDICARE IL PROPRIO NUMERO DI RIFERIMENTO:

INDICARE IL PROPRIO SESSO (M O F):

INDICARE IL PROPRIO ANNO DI NASCITA (AD ESEMPIO 1994):

INDICARE LA PROPRIA CONTRADA (SE SIETE DI UNA):

INDICARE IL PROPRIO SPORT PREFERITO:

INDICARE IL PROPRIO CANTANTE (O GRUPPO) PREFERITO:

INDICARE IL NUMERO DI RIFERIMENTO DEI PROPRI AMICI NELLA SCUOLA (MASSIMO 6):

RISPONDETE PER FAVORE ANCHE ALLE SEGUENTI DOMANDE (CON UNA CROCE SULLA RISPOSTA):

QUANTE AMICIZIE AVRANNO GLI ALTRI STUDENTI DELLA SCUOLA IN MEDIA RISPETTO A VOI?

| molte meno | un po' meno | uguali | un po' di più | molte di più |
|---|---|---|---|---|

QUANTE AMICIZIE AVRANNO I VOSTRI AMICI RISPETTO A VOI?

| molte meno | un po' meno | uguali | un po' di più | molte di più |
|---|---|---|---|---|

QUANTI DEI VOSTRI AMICI SARANNO ANCHE AMICI FRA DI LORO?

| nessuno | pochi | qualcuno | quasi tutti | tutti |
|---|---|---|---|---|

PENSATE CHE I VOSTRI AMICI ABBIANO RISPOSTO COME VOI ALLE DOMANDE DELLA PAGINA PRECEDENTE?

| nessuno | pochi | qualcuno | quasi tutti | tutti |
|---|---|---|---|---|

QUANTO TEMPO PASSATE SUI "SOCIAL NETWORKS" DI INTERNET (FACEBOOK, TWITTER...)?

| non ci vado mai | poco tempo | un po' di tempo | un bel po' | parecchio |
|---|---|---|---|---|

VI RITENETE AMANTI DEL RISCHIO?

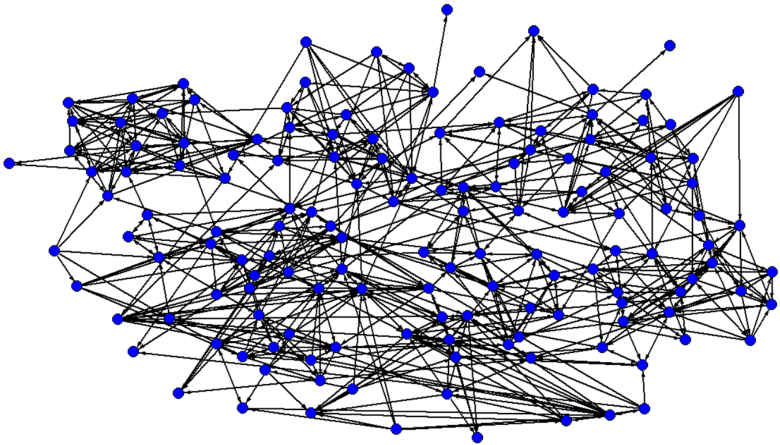
| per niente | poco | più no che si | più si che no | molto |
|---|---|---|---|---|

SE VOLETE POTETE SCRIVERE DEI COMMENTI QUI SOTTO:

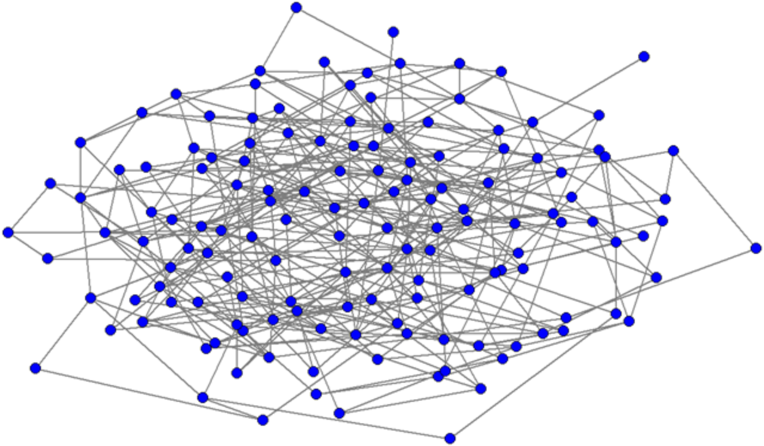
GRAZIE MILLE PER LA DISPONIBILITÀ!

UNIVERSITÀ DI SIENA 1240

# Example: the resulting directed network



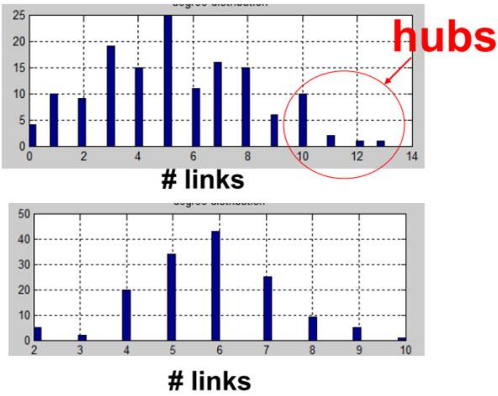
# Example: what if we *re–shuffle* links



# Example: are there differences

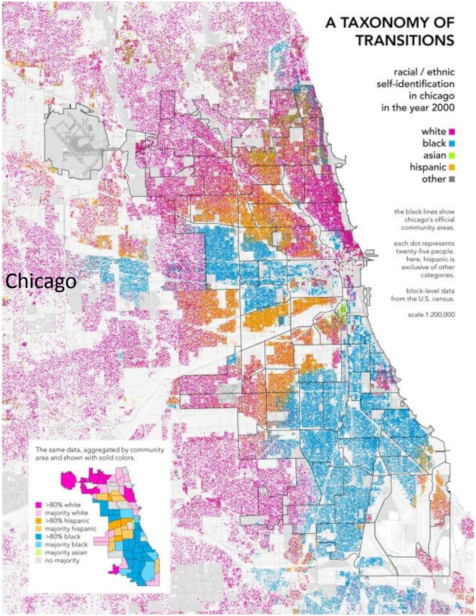
- The distribution of links is very different in the original and in randomly generated networks
- A few nodes have many links: hubs

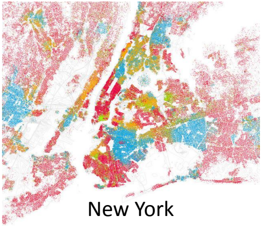
**No hubs in the random network**




ITÀ DI SIENA 1240

# Homophily


A TAXONOMY OF TRANSITIONS
racial / ethnic self-identification in chicago in the year 2000
white
black
asian
hispanic
other
the black lines show chicago's official community areas.
each dot represents twenty-five people. here, hispanic is exclusive of other categories.
block-level data from the U.S. census.
scale 1:200,000
Chicago
The same data, aggregated by community area and shown with solid colors.
>80% white
majority white
>80% hispanic
majority hispanic
>80% black
majority black
majority asian
no majority


New York

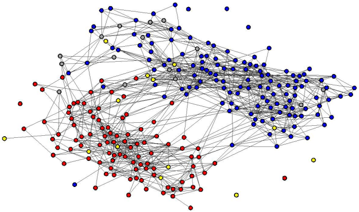

Los Angeles

.1240

# Illustrations of homophily

- National Sample: only 8% of people have any people of another race that they "discuss important matters" with (Marsden 87)
- Interracial marriages U.S.: 1% of white marriages, 5% of black marriages, 14% of Asian marriages (Fryer 07)
- In middle school, less than 10% of "expected" cross- race friendships exist (Shrum et al 88)
- Closest friend: 10% of men name a woman, 32% of women name a man (Verbrugge 77)

# Back to the school example: AddHealth and a Dutch school



| Percent: | 52 | 38 | 5 | 5 |
|---|---|---|---|---|
| | White | Black | Hispanic | Other |
| White | 86 | 7 | 47 | 74 |
| Black | 4 | 85 | 46 | 13 |
| Hispanic | 4 | 6 | 2 | 4 |
| Other | 6 | 2 | 5 | 9 |
| | 100 | 100 | 100 | 100 |

| | n=850<br>65% | n=62<br>5% | n=75<br>6% | n=100<br>7% | n=230<br>17% |
|---|---|---|---|---|---|
| | Dutch | Moroccan | Turkish | Surinamese | Other |
| Dutch | 79 | 24 | 11 | 21 | 47 |
| Moroccan | 2 | 27 | 8 | 4 | 5 |
| Turkish | 2 | 19 | 59 | 8 | 6 |
| Surinamese | 3 | 8 | 8 | 44 | 12 |
| Other | 13 | 22 | 14 | 23 | 30 |
| | 100 | 100 | 100 | 100 | 100 |

UNIVERSITÀ DI SIENA 1240

# Homophily is not contagion!

Suppose you have a dataset where people that are similar (or do similar things) are more likely to be linked together. . .

is it homophily or contagion/learning/imitation?

To answer this question is an incredibly difficult task!
Probably reasonable answers can only be obtained from controlled experiments. . .

We will not address this issue here, and we will stick to characteristics that are (or can be thought of) as **exogenous**

# Imbreeding Homophily

Type $i$ has a representativeness $w_i \in [0,1]$ in the population

On average a type $i$ student has a ratio $q_i \in [0,1]$ of same–type friends

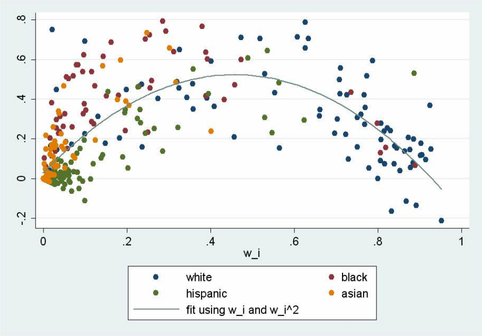
Coleman (1958) defines an index of inbreeding homophily of group $i$:

$$IH_i = \frac{q_i - w_i}{1 - w_i}.$$

This measure is normalized in $[0,1]$ whenever there is a non–negative level of homophily (i.e. $q_i \geq w_i$)

# Imbreeding Homophily in the Add-Health data

Each point is one racial group in one school




Why do we have a bell shape (consistent in other datasets)?

This was an open problem in sociology

UNIVERSITÀ DI SIENA 1240

# Reasons for homophily

Idea of reduced form

- communication costs
- frictions
- opportunities
- choices

It is important to distinguish between choices and opportunities (we'll see. . . )

# Strength of Weak Ties

Granovetter (1973) interviews: 54 people who found their jobs via social tie:

- 16.7 percent via strong tie (at least two interactions/week)
- 55.7 percent via medium tie (at least one interaction per year)
- 27.6 percent via a weak tie (less than one interaction per year)

Theory: weak ties form 'bridges', less redundant information

It seems in contrast with homophily but is actually due to it

Structural holes (Burt, 2004):

- a few nodes in the society convey all the information
- if they are removed information flow is broken
- they are *strong* nodes with many *weak* ties

UNIVERSITÀ DI SIENA 1240

# An economic model of friendship

# An Economic Model of Friendship: Homophily, Minorities and Segregation (ECMA, 2009)

- Most networks involve both choice and chance in formation
- What are the relative roles?
- Random/Strategic models can be too extreme
- Can we see relative roles in homophily?

Group A and Group B form fewer cross race friendships than would be expected given population mix

- Is it due to structure: few meetings?
- Is it due to preferences of group A?
- Is it due to preferences of group B?

Revealed preference theory

- Common to Consumer Theory
- Use it in mapping social/friendship choices too!
- Different information than surveys on racial attitudes

- Utilities specified as a function of friendships
- Meeting process that incorporates randomness
- Allow both utilities and meeting process to depend on types

- Types: $i \in \{1, \ldots, K\}$
- $s_i$ is # of same type friends
- $d_i$ is # of different type friends

$$U_i = (s : i + \gamma_i d_i)^\alpha$$

- utility to type $i$: $\gamma_i$ is preference bias – while $\alpha < 1$ captures diminishing returns

UNIVERSITÀ DI SIENA 1240

# Individual choice

- $t_i$ number of friends – proportional to time spent socializing
- $q_i$ fraction of friends that will be of own type
- $t_i$ maximizes

$$(q_i t_i + \gamma_i (1 - q_i) t_i)^{\alpha} - c t_i$$

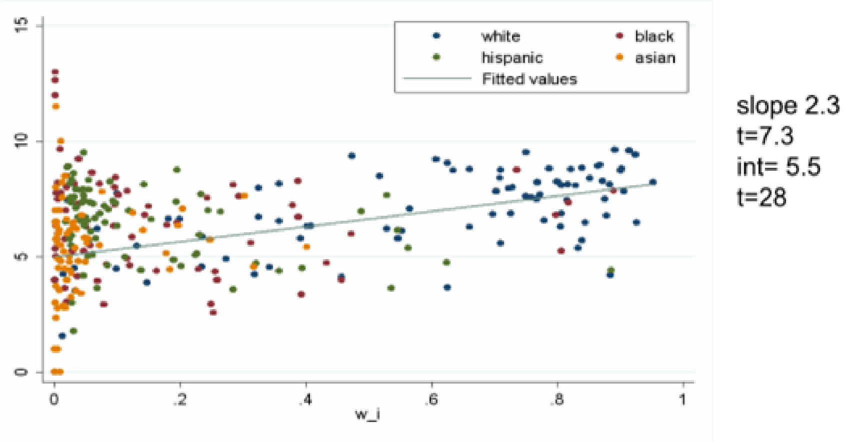
- Solution

$$t_i = (\alpha / c)^{1/(1-\alpha} (q_i + \gamma_i (1 - q_i))^{\alpha/(1-\alpha)}$$

- If $\gamma_i < 1$ then this is increasing in $q_i$

# Friends increase with $q_i$

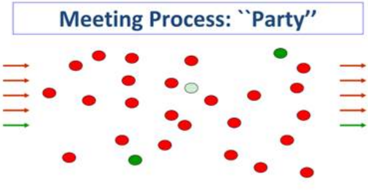
Add Health data – $w_i$ is the size of groups

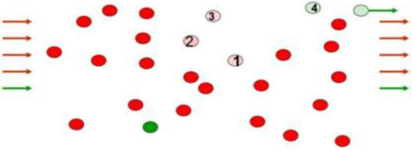



Group size (fraction)

# Where do $q_i$'s come from?

Randomness in meetings, but also have $q_i$'s determined by the decisions of the agents



# Bias in Meeting Process

- $t_i$ number of friends – proportional to time spent socializing
- $q_i$ rate at which type $i$ meets type $i$
- $1 - q_i$ rate at which type $i$ meets other types

$$q_i = (\text{stock in the pool})^{1/\beta_i}$$

- $\beta_i > 1$ meet own types faster than stocks (matching technology)
- $stock_i = \frac{w_i t_i}{\sum_j w_j t_j} = q_i^{\beta_i}$
- balance condition: $\sum_j q_j^{\beta_j} = 1$
- atomless population (individual has no effect on proportions)

In equilibrium, we put together the choice equations with the balance condition
We obtain that when both are present. . .

UNIVERSITÀ DI SIENA 1240

. . . we match very well the empirics

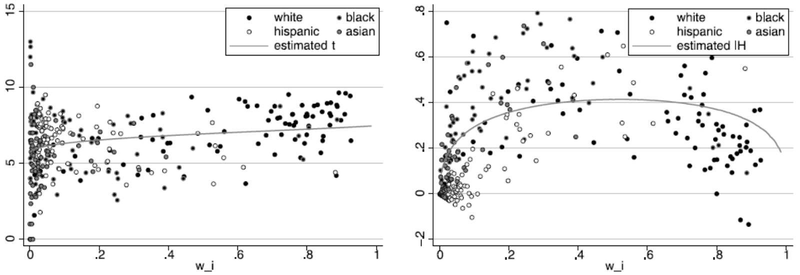



FIGURE 6.—Fitted model and data: left, inbreeding homophily by relative group size; right, number of friends by relative group size. Parameter values are $\alpha = 0.2$, $\beta = 2$, $\gamma = 0.8$, and $c = 0.04$.

The bell–shape was attributed to conflicting coalitions,
we have an explanation based on individual opportunities and choices

UNIVERSITÀ DI SIENA 1240

# Learning in a homophilous network

# When agents are sophisticated... things are complicated

- Bayesian approach is computationally demanding in network settings
- It *usually* gives consensus (herding effect)

Bala & Goyal (1998)

- $n$ players in an undirected component g
- Choose action A or B each period
- $A$ pays 1 for sure, $B$ pays 2 with probability $p$ and 0 with probability $1-p$
- Each period get a payoff based on choice – Also observe neighbors' choices
- Maximize discounted stream of payoffs $E\left(\sum_t \delta^t \pi_{it}\right)$

UNIVERSITÀ DI SIENA 1240

# Result in Bala & Goyal (1998)

If $p$ is not exactly $1/2$, then with probability 1 there is a time such that all agents in a given component play just one action (and all play the same action) from that time onward

**Sketch of proof**

- Suppose contrary: some play A and some play B
- Some agent in some component plays B infinitely often
- That agent will converge to true belief by the law of large numbers
- Must be that belief converges to $p > 1/2$, or that agent would stop playing B
- With probability 1, all agents who see B played infinitely often converge to a belief that B pays 2 with prob $p > 1/2$
- All agents will end up playing B

**Herding:** they may as well all converge to the wrong action $A$ (when $p > 1/2$)!

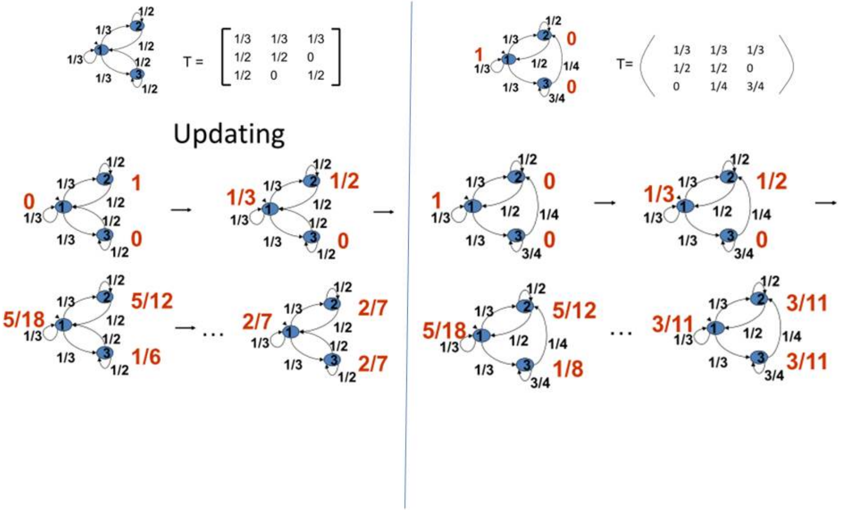
# Naive learning

- Repeated communication
- Information comes only once
- See how information disseminates
- Who has influence, convergence speed, network structure impact...
- Repeatedly average beliefs of self with neighbors
- Non–Bayesian if weights do not adjust over time

- Individuals $\{1, 2, \ldots, n\}$
- $T$ weighted directed network (adjacency matrix) – it is a stochastic matrix
- Start with beliefs, attitude, etc. $b_i(0) \in [0, 1]$ – but also vectors work
- Updating: $b_i(t) = \sum_j T_{ij} b_j(t-1)$

UNIVERSITÀ DI SIENA 1240

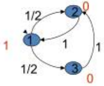
# Examples of convergence




SIENA 1240

# Does it always converge?

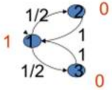
## Example - Convergence:




$$T=\begin{pmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

$$b(0)=\begin{pmatrix}1\\0\\0\end{pmatrix} \quad b(1)=\begin{pmatrix}0\\1\\0\end{pmatrix} \rightarrow \begin{pmatrix}1/2\\0\\1\end{pmatrix} \rightarrow \begin{pmatrix}3/4\\1/2\\0\end{pmatrix} \rightarrow \begin{pmatrix}1/4\\3/4\\1/2\end{pmatrix} \ldots \rightarrow \begin{pmatrix}2/5\\2/5\\2/5\end{pmatrix}$$

## Example – No Convergence:




$$T=\begin{pmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

$$b(0)=\begin{pmatrix}1\\0\\0\end{pmatrix} \quad b(1)=\begin{pmatrix}0\\1\\1\end{pmatrix} \rightarrow \begin{pmatrix}1\\0\\0\end{pmatrix} \rightarrow \begin{pmatrix}0\\1\\1\end{pmatrix} \rightarrow \begin{pmatrix}1\\0\\0\end{pmatrix} \ldots \rightarrow$$

.1240

# Theory

- $T$ converges if $\lim T^t \vec{b}$ exists for all $\vec{b}$
- $T$ is *aperiodic* if the greatest common divisor of its cycle lengths is one


**Result:** $T$ converges only if it is (strongly) connected and aperiodic

Why do we need 'connected'?

Aperiodicity is easy to satisfy: just a little weight on own past opinions for one agent

# Who has influence?

- When group reaches a consensus, what is it?
- Who are the influential agents in terms of shaping the limiting belief?

## Influence:

$$\textbf{Limit} \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 1/2 & 0 \\ 0 & 1/4 & 3/4 \end{pmatrix}^{t} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 3/11 \\ 3/11 \\ 3/11 \end{pmatrix}$$

$$\textbf{Limit} \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 1/2 & 0 \\ 0 & 1/4 & 3/4 \end{pmatrix}^{t} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 4/11 \\ 4/11 \\ 4/11 \end{pmatrix}$$

$$\textbf{Limit} \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 1/2 & 0 \\ 0 & 1/4 & 3/4 \end{pmatrix}^{t} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 4/11 \\ 4/11 \\ 4/11 \end{pmatrix}$$

# Who has influence?

- What do rows of $T^t$ converge to?
- Look for a row vector $\vec{s}$ indicating the relative influence each agent has – limit belief is $\vec{s} \cdot \vec{b}$
- We must have $\vec{s} \cdot \vec{b} = \vec{s} T \vec{b}$
- So, $\vec{s} = \vec{s} T$ : $\vec{s}$ is the left unit eigenvector

- $s_i = \sum_j T_{ji} s_j$
- Recursive definition: High influence from being paid attention to by people with high influence...
- Base idea of *Google page rank*

UNIVERSITÀ DI SIENA 1240

# First implications

Stubborn agents

- An agent who places high weight on self will maintain belief while others converge to that agent's belief
- Groups that are highly introspective will have substantial influence.

It is based on an *eigenvector centrality*

- it provides foundation for eigenvector-based centrality or power measures

No *gurus* or stubborn agents

- the network will average the information: "the wisdom of crowd"
- no *herding* as could happen in a Bayesian model

UNIVERSITÀ DI SIENA 1240

# How homophily affects the speed of learning and best–response duynamics (QJE, 2012)

The *naïve* model approximates well real world behavior

- this assumption is based on models with rational Bayesian agents
- but in lab experiments it does so even better than Bayesian models

Even if convergence of opinions is reached in the long–run,
it may take a lot of time

Example:

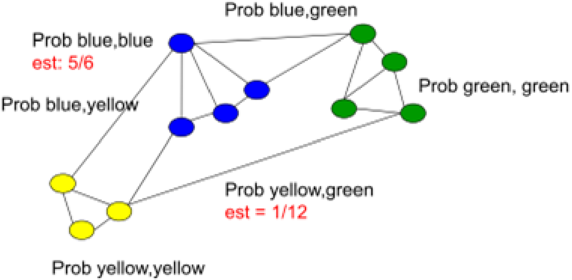
- Iraq was invaded in March 2003
- in October 2004, 47% of Republicans and 9% of Democrats believed that Iraq had weapons of mass destruction
- in March 2006, those probabilities fell only to 41% and 7%

UNIVERSITÀ DI SIENA 1240

# Simple model of network with homophily

Probabilities depend on characteristics
Extend the basic Erdos-Renyi $G(n, p)$ model:

Nodes have characteristics: e.g., age, gender, religion, profession, etc.
links between nodes depend on the pairs' characteristics




Reminds Granovetter's story

UNIVERSITÀ DI SIENA 1240

# Multi–type random networks

- Assume $n$ nodes partitioned in $K$ exogenous types
- Type $k$ has $n_k$ nodes (this is represented by a $K$–dimensional vector $\vec{n}$)
- The probability of linking between a type–$i$ and a type–$j$ is given by $p_{ij}$
- Probabilities are summarized by the $K \times K$ matrix $\hat{P}$
- This describes a class of random matrices that we call $\hat{A}(\hat{P}, \vec{n})$

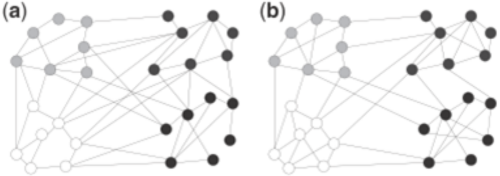



FIGURE I

Islands networks with low and high homophily are shown in (a) and (b), respectively. Nodes that are shaded differently are of distinct types.

UNIVERSITÀ DI SIENA 1240

# Spectral homophily

We define a measure of homophily in multi–type random networks

First of all take expectations from $\hat{P}$:

$$F_{i,j} = \frac{n_i \cdot n_j \cdot p_{ij}}{\sum_k n_i \cdot n_k \cdot p_{ik}}$$

$\hat{F}$ is the expected fraction of links that nodes of one type will have with nodes of other types

It is possible to express the expected evolution of opinions in the DeGroot model only with powers $\hat{F}$

Definition (Spectral homophily)

$h^{spec}(\hat{P}, \vec{n})$ is the second largest eigenvalue of $\hat{F}$.

**Intuition**: it measures how *easily* you can break the network in two disconnected components

UNIVERSITÀ DI SIENA 1240

# Islands example

- There are $m$ *islands*
- Suppose that $\hat{P}$ has values $p_s$ on the diagonal and $p_d$ between different types
- $p = \frac{p_s+(m-1)p_d}{m}$ is the common expected degree

We have

$$h^{spec}(\hat{P}, \vec{n}) = \frac{p_s - p_d}{mp} = \frac{\frac{p_s}{mp} - \frac{1}{m}}{1 - \frac{1}{m}} = \frac{q_i - w_i}{1 - w_i} = IH_i$$

In this symmetric case, spectral homophily coincides with Coleman index!

UNIVERSITÀ DI SIENA 1240

# Consensus time

Now let's go back to the DeGroot learning model

We can measure how fast a consensus is reached in a connected network $\hat{A}$

**Definition (Consensus time)**

$$CT(\epsilon, \hat{A}) \sup_{\vec{b}} \min\{t : ||\hat{A}^t \vec{b} - \hat{A}^\infty \vec{b}|| < \epsilon\}$$

It is the time needed to reach *enough* consensus.

It is defined on a worst–case scenario for $\vec{b}$:
what is the worst disposition of initial beliefs, so that you need at least $t$ time steps before...

UNIVERSITÀ DI SIENA 1240

# Consensus time depends only on homophily

For multi-type random networks we have a nice result

**Proposition**

*For any* $\gamma > 0$

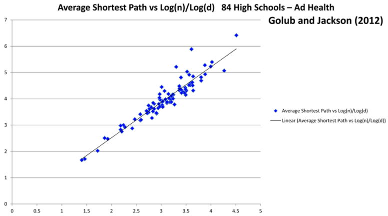
$$CT\left(\gamma/n, \hat{A}(\hat{P}, \vec{n})\right) \propto \frac{\log(n)}{\log\left(1/|h^{spec}(\hat{P}, \vec{n})|\right)}$$

Independently on the required precision, the consensus time will depend only on the size of the population, and on the homophily of the network

UNIVERSITÀ DI SIENA 1240

# Comparison with contagion

On the other hand, the diameter of the network depends only on its average expected degree $d$




Naive learning

- Homophily slows down convergence
- Average density does not matter

Simple contagion model (SI)

- Average density increases the contagion process
- Homophily does not matter

# Voting example

There is a nice voting example in Golub and Jackson (2012)

We know that the limit of the belief propagation converges to the common average,
but in–between closed groups may temporarily diverge

Consider a majority of relative size $M > 1/2$, a true state of nature $\omega \in \{0,1\}$ and the following random signals, also from $\{0,1\}$:

- a member of the majority observes the true signal with probability $\mu < 1/2$
- a member of the minority observes the true signal with probability $\nu > \mu$
- we have also $p = M\mu + (1-M)\nu > 1/2$

If they had to vote for the correct state in $\{0,1\}$, without communication, they would (in expectation) vote correctly

# Voting example with communication

Suppose that there is a network of communication based on:

$$\hat{F} = \begin{pmatrix} 1-f & f \\ f\frac{M}{1-M} & 1-f\frac{M}{1-M} \end{pmatrix}$$

with high expected degree and no weight to previous individual opinion

If $f < 1 - M$ we have homophily

In this case

$$h^{spec}(\hat{P}, \vec{n}) = 1 - \frac{f}{1-M}$$

both $f$ and $M$ increase speed of convergence

Also if the nodes could communicate an infinite amount of time they would vote correctly, because the common opinion would converge to $p$ (the wisdom of croud)

# Voting example with communication

What happens for limited communication?

Suppose $\omega = 1$ and $f < 1 - M$.
After $t > 0$ rounds of communication the expected beliefs of the two groups will be respectively

$$\begin{aligned} b_{maj} &= p - \left(1 - \frac{f}{1-M}\right)^t (p - \mu) < p \\ b_{min} &= p + \frac{M}{1-M}\left(1 - \frac{f}{1-M}\right)^t (p - \mu) > p \end{aligned}$$

As long as $p - \left(1 - \frac{f}{1-M}\right)^t (p - \mu) < \frac{1}{2}$ all members of the majority will vote wrong!

So, if $\left(1 - \frac{f}{1-M}\right)^t > \frac{p-1/2}{p-\mu}$, the population will not vote correctly anymore!

# A job market model

# The Effects of Social Networks on Employment and Inequality (AER, 2004)

Consider $N$ agents on an exogenous undirected network

- time is discrete and infinite
- they are workers, with employment status $s_{i,t} \in \{0, 1\}$
- job opportunities are exogenous
- they may hear about a job with i.i.d. probability $a$ (formal job market)
- if already employed, they pass the job to an unemployed neighbor (informal job market)
- they can be fired with i.i.d. probability $b$

This is an ergodic markov process where each state is described by a vector $\vec{s}_t \in \{0, 1\}^n$

UNIVERSITÀ DI SIENA 1240

# One–lag correlation

In one time–step, you benefit from your neighbors being employed

And you compete with the neighbors of your neighbors for getting second–hand job info

You don't care about more distant nodes

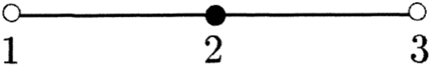



FIGURE 1. NEGATIVE CORRELATION IN CONDITIONAL EMPLOYMENT

What happens in the long run? Do I still have some correlation with neighbors?

UNIVERSITÀ DI SIENA 1240

# Position in the network

### Proposition

*The unique steady–state long–run distribution on employment is such that the employment statuses of any path–connected agents are positively correlated*

Example with $a = 0.1$ and $b = 0.015$

| $g$ | $\text{Prob}(s_1 = 0)$ | $\text{Corr}(s_1, s_2)$ | $\text{Corr}(s_1, s_3)$ |
|---|---|---|---|
|  | 0.132 | – | – |
|  | 0.083 | 0.041 | – |
|  | 0.063 | 0.025 | 0.019 |
|  | 0.050 | 0.025 | 0.025 |

FIGURE 2. CORRELATION AND NETWORK STRUCTURE I

# Long–run employment probabilities

But still the network position matters!

Example with $a = 0.1$ and $b = 0.015$ (as before)

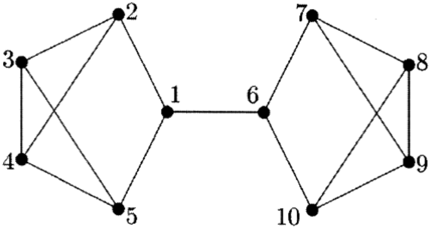



FIGURE 4. A NETWORK WITH A BRIDGE

They have all the same degree. Remember *structural holes*?

In this case, nodes 1 and 6 have unemployemnt probability 4.7%,
nodes 2, 5, 7 and 10: probability 4.8%,
nodes 3, 4, 8 and 9: 5%.

# Persistence in unemployment

## Proposition

*Starting under the steady–state distribution, the conditional probability that an individual will become employed in a given period is decreasing with the length of her observed (individual) unemployment spell.*

| $g$ | 1 period | 2 periods | 10 periods | limit |
|---|---|---|---|---|
|  | 0.099 | 0.099 | 0.099 | 0.099 |
|  | 0.176 | 0.175 | 0.170 | 0.099 |
|  | 0.305 | 0.300 | 0.278 | 0.099 |

FIGURE 6. DURATION DEPENDENCE

- Being employed is a local public good (a case of strategic substitutes),
- but agents cannot cordinate efficiently (it is not a choice)
- a region of the social network may temporarily suffer diffused unemployment

UNIVERSITÀ DI SIENA 1240

# Drop out decision – A case of strategic complements

The model with an exogenous network has long–run positive correlation. But what happens if agents can drop out from the labor market?

Labor Participation Decisions (Calvo–Armengol & Jackson 04,07,09)

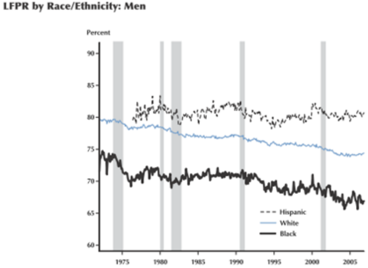
- Value to being in the labor market depends on number of friends in labor force
- Drop out if some number of friends drop out
- Participate if at least some fraction of friends do
- Some heterogeneity in threshold (different costs, natural abilities...)
- Homophily – segregation in network
- Different starting conditions: history...

UNIVERSITÀ DI SIENA 1240

# Real data

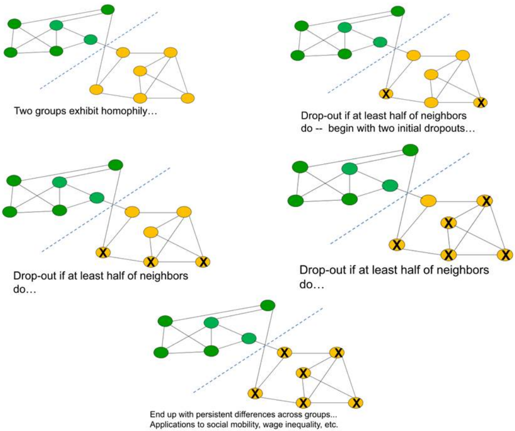
Drop-Out Rates & Labor Force Participation Rate
(Chandra (2000), DiCecio et al (2008) – males 25 to 55)

| | 1940 | 1950 | 1960 | 1970 | 1980 | 1990 |
|---|---|---|---|---|---|---|
| whites | 3.3 | 4.2 | 3.0 | 3.5 | 4.8 | 4.9 |
| blacks | 4.2 | 7.5 | 6.9 | 8.9 | 12.7 | 12.7 |




UNIVERSITÀ DI SIENA 1240

# Example with homophily




UNIVERSITÀ DI SIENA 1240

# Take–home message

# Networks and homophily

In many contexts it is important to study the constraints induced by social networks

- communication
- diffusion
- market opportunities

Homophily is a common feature of social networks

- because of individual choices
- and because of meeting biases
- the two effects reinforce each others

Homophily affects temporary of even persisting inequalities in networks

- in the updating of opinions
- in the sharing of goods
- in coordination problems