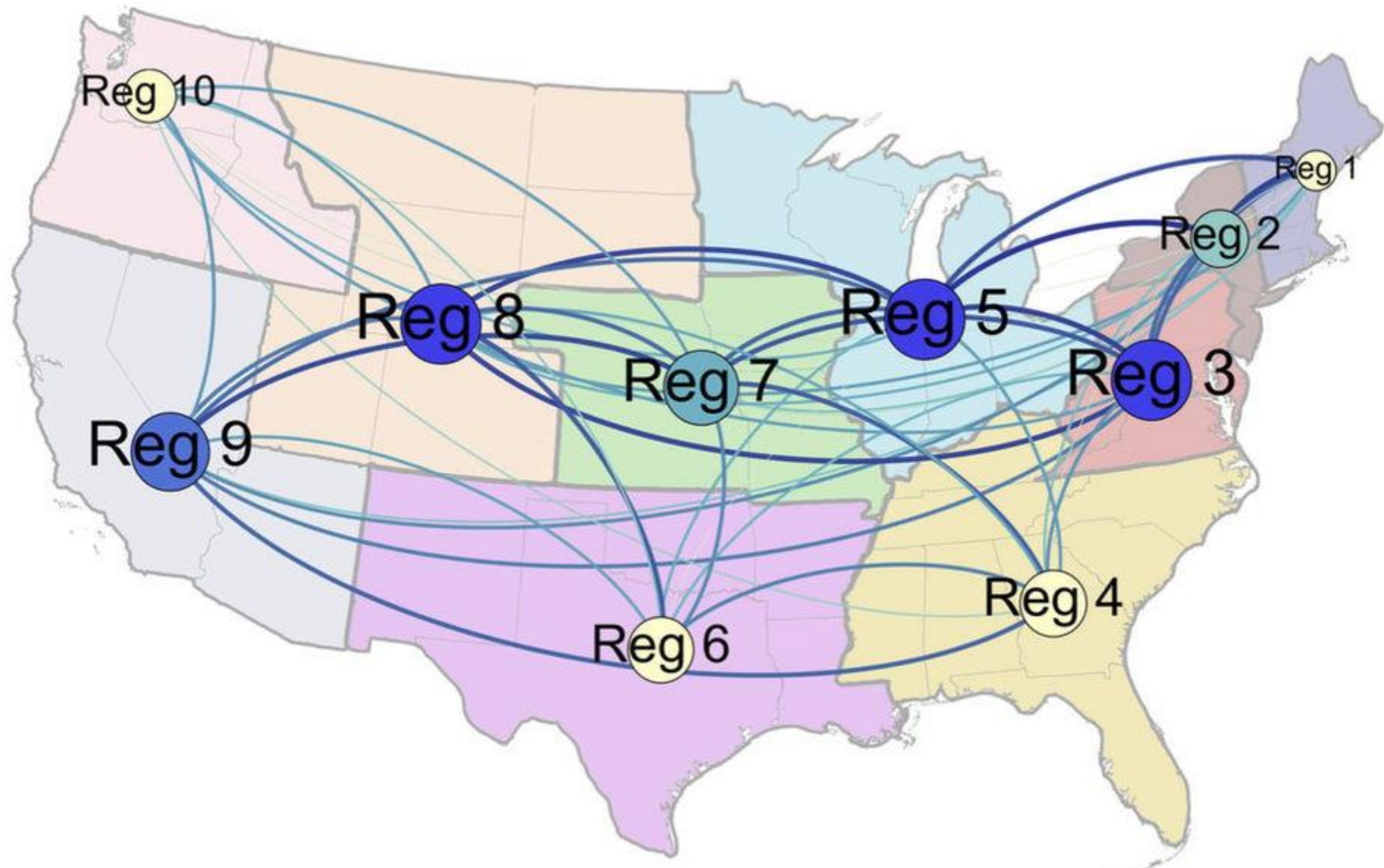


Как не попасться в сети?

Клачкова Ольга,
экономический факультет МГУ имени М.В.Ломоносова

Plague Inc., 2012





Контроль инфекционных заболеваний

[[nature.com/scientific reports](https://www.nature.com/scientific-reports)]

США, сезонный грипп (каждый год новый вирус...), 5-20% населения ежегодно

U.S. Centers for Disease Control and Prevention – исторические корреляции на основе лабораторно подтвержденных случаев – лаг 1-2 недели

Google Flu Trends – поисковые запросы – лаг в 1-2 дня

- How densely connected does a society have to be in order to have an infection reach a nontrivial fraction of its members?
- How does it depend on the particulars of the social network in place?
- How does this depend on the infectiousness of the disease?

Социальные и экономические сети – графическая визуализация взаимодействий между различными агентами:

- индивиды;
- фирмы;
- ...

Типичная форма представления – граф.

Что такое граф?

Граф $G = (V, E)$ – пара двух множеств:

V – множество вершин (vertex)

$E = \{e_{ab} = (a, b): a, b \in V\}$ – множество пар вершин – дуг, ребер (edge)

Последовательность дуг графа – путь, если начало каждой последующей дуги совпадает с концом предыдущей.

Нагруженный граф: каждой дуге $e \in E$ поставлено в соответствие число $w(a, b)$ (вес дуги).

Длина пути – сумма весов входящих в путь дуг.

Числовые характеристики графа



Компонента связности это подграф, у которого все вершины имеют путь до любой другой вершины компоненты.

Степень вершины – количество рёбер, примыкающих к вершине.

Кратчайший путь между парой вершин – путь минимальной длины между парой вершин.

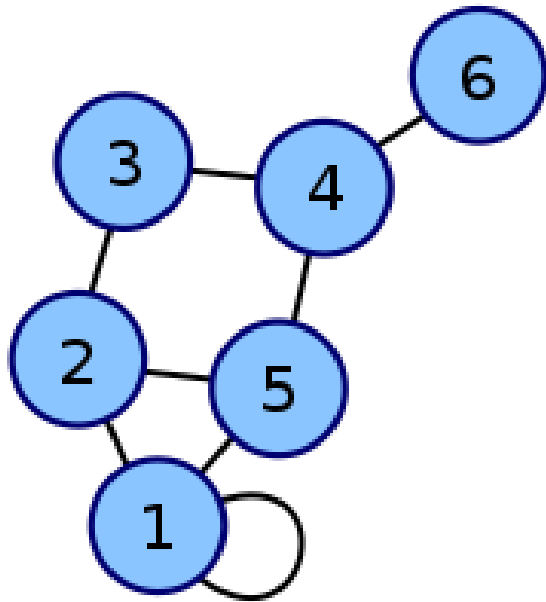
Facebook – средняя длина кратчайшего пути = 4,57 (теория 6-ти рукопожатий).

Матрица смежности

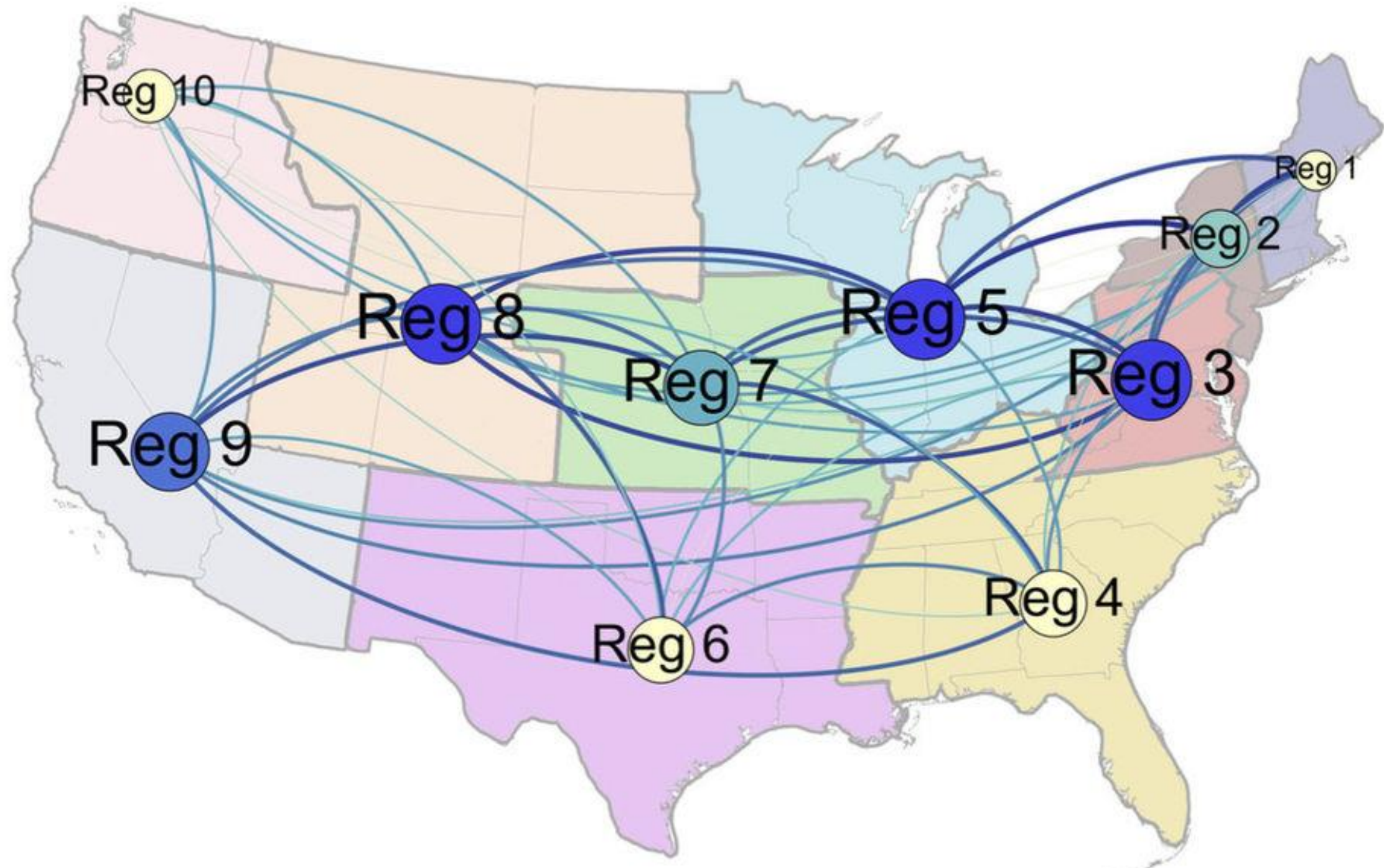
Дан неориентированный граф $G = (V, E)$.

Пусть множество вершин $V = \{1, 2, \dots, n\}$.

Матрица размера $n \times n$: $M = \{m_{ij}\} = \begin{cases} 1, & \text{if } (i, j) \in E \\ 0, & \text{if } (i, j) \notin E \end{cases}$



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



Характеристики графа: «важность» вершины

Centrality = важность, центральность.

Центральность графа:

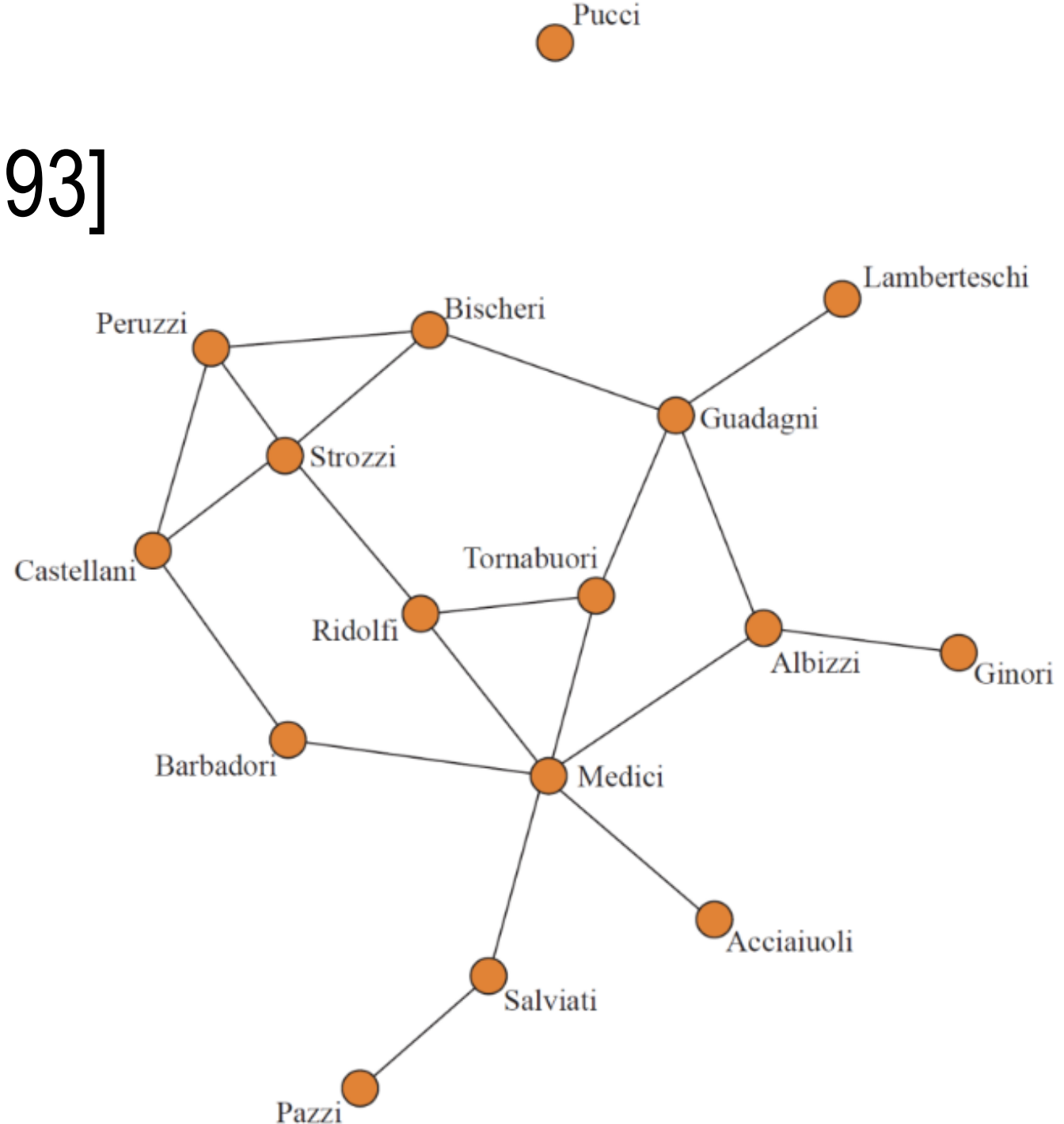
- по степени вершины (degree centrality);
- по кратчайшему пути (betweenness centrality);
- по собственному значению (eigenvalue centrality).

Исторический пример [Padgett and Ansell, 1993]

Вершины – итальянские аристократические семьи Флоренции в начале XV века.

Дуга: две семьи соединены отрезком, если они связаны друг с другом посредством брачного союза.

Какая из семей завоевала власть в городе и почему?



Сетевые «причины» могущества

Коэффициент центральности (betweenness centrality) вершины k :

$$P_k = \frac{1}{(n-1)(n-2)/2} \cdot \sum_{\substack{i,j;i \neq j, \\ k \notin \{i,j\}}}^n \frac{P_k(i,j)}{P(i,j)}$$

$P(i, j)$ – число кратчайших путей между вершинами i и j

$P_k(i, j)$ – число кратчайших путей между вершинами i и j , проходящих через вершину k

Сетевые «причины» могущества

⇒ Коэффициент центральности – степень влияния одной семьи над другими.

Коэффициент центральности семьи Медичи равен 0.522: Медичи лежат на более чем половине кратчайших путей между всеми другими семьями!

Ближайший конкурент Медичи - семья Гаданьи, и её коэффициент центральности равен 0.255.

В среднем, этот показатель равен 0.074.

The Diffusion of Microfinance, Banerjee, Duflo, Jackson, 2013

Как распространять полезную информацию в сети, чтобы наибольшее число людей могло ей воспользоваться?

- сетевые характеристики первого индивида, получающего информацию;
- индивид более вероятно узнает о продукте, если большее число его друзей может ему о нем рассказать;
- решение об использовании продукта информированным индивидом зависит от аналогичных решений его друзей.

75 деревень, Карнатака, юго-западная Индия, ≥ 46 км между деревнями, похожие по демографическим показателям

Bharatha Swamukti Samsthe (BSS) – микрофинансовая организация

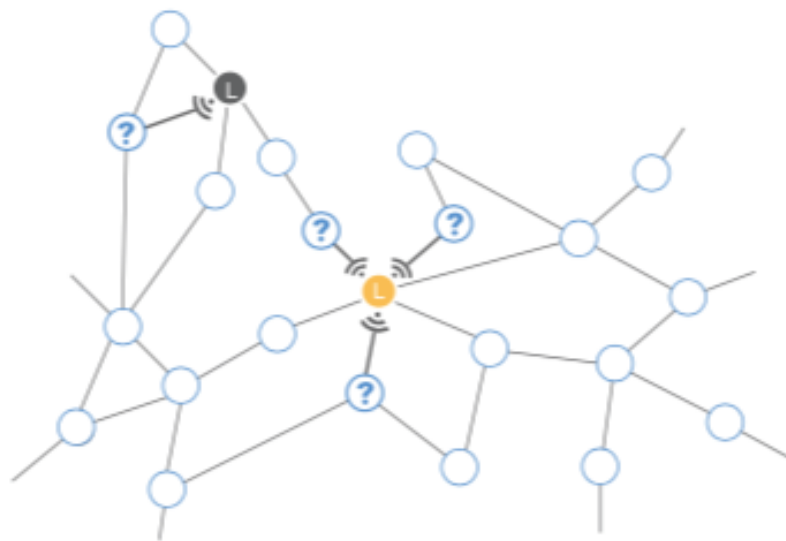
Social network data along 12 dimensions:

- names of those who visit the respondent's home,
- those whose homes the respondent visits,
- kin in the village,
- nonrelatives with whom the respondent socializes,
- those from whom the respondent receives medical advice,
- those from whom the respondent would borrow money, those to whom the respondent would lend money, those from whom the respondent would borrow material goods (kerosene, rice, etc.), those to whom the respondent would lend material goods,
- those from whom the respondent gets advice, those to whom the respondent gives advice,
- those with whom the respondent goes to pray (at a temple, church, or mosque)

A Leaders are informed and make a decision on participation.



B Information is passed on by leaders; leadership participation affects probability of information sharing.



C Newly informed nodes make a decision on participation.



Fig. 1. Diffusion of information and participation. (A) Informed leaders. (B) Leaders passing information; the participating leader passes at a higher rate. (C) The participation decision of newly informed nodes. (D) Nodes pass information again, which can vary according to participation. (E) Newly informed nodes deciding again to participate.

D All informed nodes pass on information further; the probability of information sharing is, again, based on participation.



E Fresh round of newly informed nodes make participation decision.



q_P – вероятность получить информацию, если рассказывает «сосед», решивший использовать информацию

q_N – вероятность получить информацию, если рассказывает «сосед», решивший НЕ использовать информацию

Оценка q_N , q_P симуляцией по сети по соответствию реальным параметрам:

- share of leaders who participate in microfinance;
- share of households with no participating neighbors that participate;
- share of households in the neighborhood of a participating leader that participate;
- share of households in the neighborhood of a nonparticipating leader that participate.

| q^N | q^P |
|--------------|---------------|
| 0.050 (0.01) | 0.350 (0.124) |

- q_P и q_N значительно различаются;
- важна передача информации НЕучастниками;
- нет особой роли лидера (!) в смысле вероятностей.

Как же выбрать лидера?

У нас есть оценки q_N , q_P и социальная сеть \rightarrow поставим в соответствие i -ой вершине долю домохозяйств, которые стали бы участниками, если бы i -ая вершина была лидером (единственной проинформированной)

\rightarrow communication centrality

Эмпирически, communication centrality значимо и положительно влияет на долю домохозяйств, ставших участниками.

Корреляция с остальными мерами центральности не значима.

Диффузия чего моделируется?

- **Принятие новых технологий и продуктов (модель Басса)**
Tucker, 2008: распространение новой технологии видео-сообщений в банке с 2118 сотрудниками; большой эффект, если новую технологию использует или формальный начальник, или неформальный лидер.
- **Политические идеи**
Huckfeldt, Sprague, 1995: South Bend, Indiana, выборы 1984; большая вероятность, что «соседи» по графу имеют схожее политическое мировоззрение.
- **Финансовая неустойчивость**
Elliott, Golub, Jackson, 2014: вероятность банкротства компании A зависит от:
степени интеграции: какая доля активов компании A находится в собственности других компаний?
степени диверсификации: в собственности скольких других компаний находятся активы компании A ?

Диффузия чего моделируется?

Всего, что душа пожелает, были бы данные:

- Christakis, Fowler, 2007: obesity levels – 12067 человек в США с 1971 по 2003 год; a person's chances of becoming obese increased by 57% if he or she had a friend who became obese.
- Glaeser, Sacerdote, Scheinkman, 1996: cross-city variance in crime rates in the U.S.
- Bearman, Moody, Stovel, 2004: romantic connections in high-school.

Неравенство и характеристики сети: рынок труда

Формальный vs. неформальный поиск работы:

- 15% безработных в 1970 г. ищет работу через друзей и родственников, 23% в 1992 (Bradshaw, 1973; Ports, 1993);
- доли варьируются в зависимости от соц.-дем. характеристик индивидов;
- 30-60% людей работают по рекомендациям друзей и родственников (Bewley, 1999) → реже уходят и дольше работают, а про зарплаты – неясно (стартовая выше, а дальше рост медленнее).

Связь неформального поиска работы и неравенства в зарплатах и занятости на рынке труда?

⇒ Calvo-Armegnol, Jackson (2002; 2003)

Знаем структуру существующего социального графа и
изучаем его характеристики

vs.

Моделируем формирование структуры социального графа

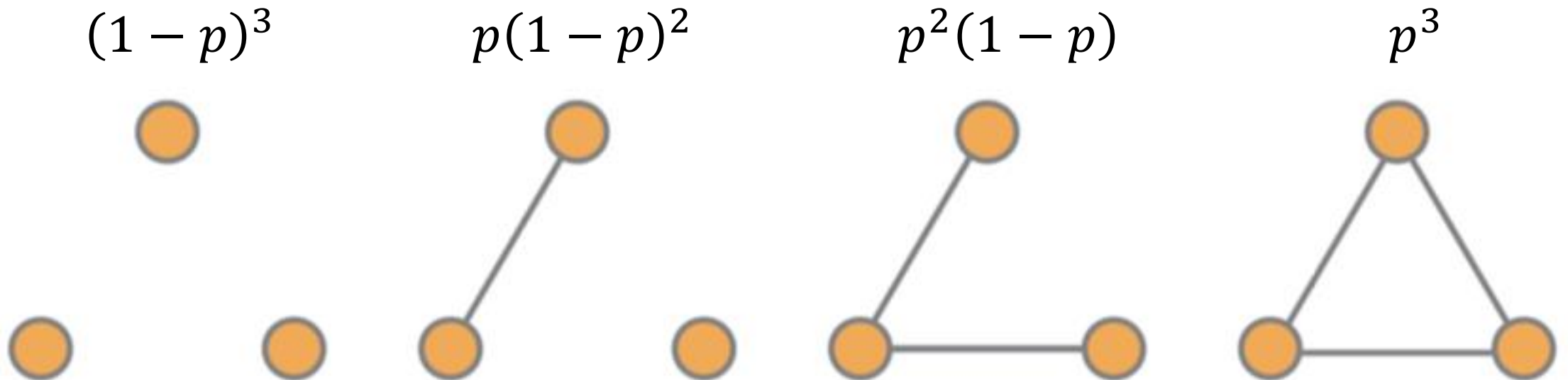
Случайные графы [Erdős, Rényi, 1959-1961]

Пусть множество вершин $V = \{1, 2, \dots, n\}$.

Вероятность появления ребра (i, j) равна $0 < p < 1$.

Появление ребра – независимо от других ребер.

⇒ Биномиальная модель.



Случайные графы [Erdős, Rényi, 1959-1961]

Для любого случайного графа вероятность того, что у него будет m ребер при n вершинах:

$$p^m \cdot (1 - p)^{\frac{n(n-1)}{2} - m}$$

Степень вершины в случайном графе (вероятность того, что вершина будет иметь степень $= d$):

$$C_{n-1}^d \cdot p^d \cdot (1 - p)^{n-1-d} \quad \rightarrow \quad ???$$

$n \rightarrow \infty$
 $p \rightarrow 0$

Случайные графы [Erdős, Rényi, 1959-1961]

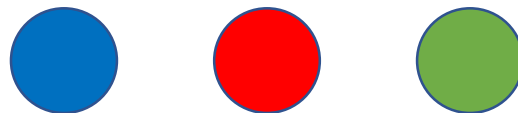
Степень вершины в случайном графе (вероятность того, что вершина будет иметь степень = d):

$$C_{n-1}^d \cdot p^d \cdot (1-p)^{n-1-d} \quad \rightarrow \quad e^{-(n-1)p} \frac{((n-1)p)^d}{d!}$$

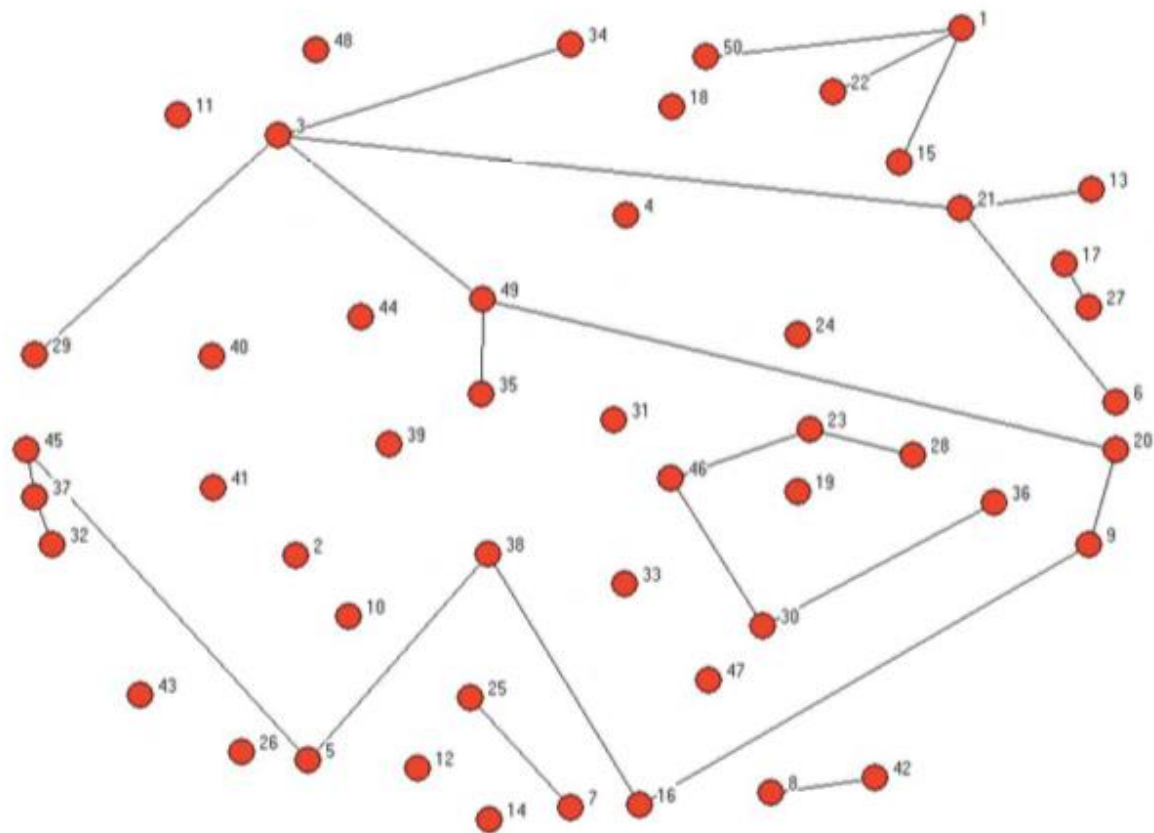
$n \rightarrow \infty$
 $p \rightarrow 0$

Пуассон!

В пределе, при $n \rightarrow \infty$, совпадает с долей вершин, имеющих степень d .



Случайные графы [Erdős, Rényi, 1959-1961]

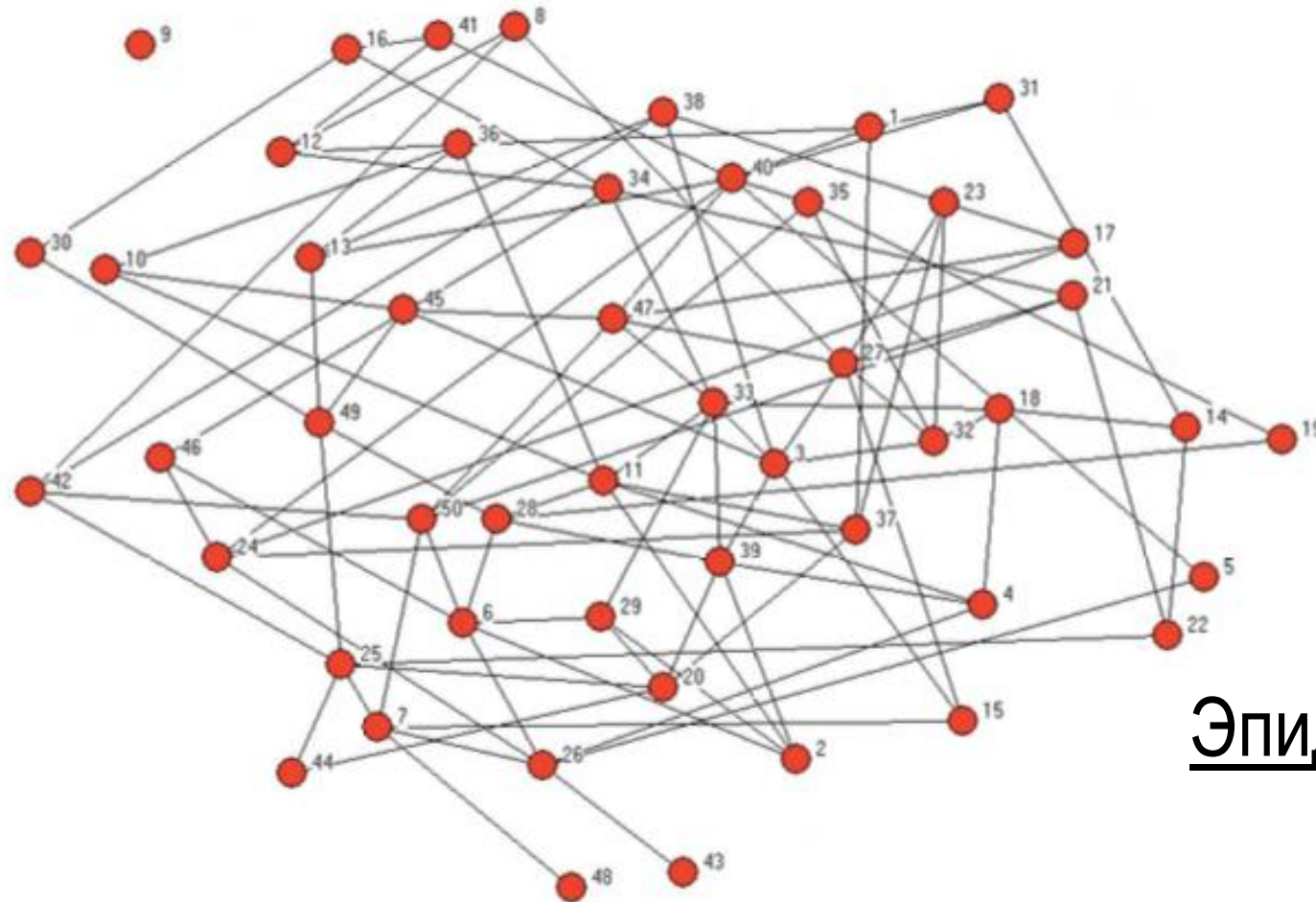


$$n = 50, p = \frac{1}{50} = 0.02$$

$$P(d = 0) = e^{-(n-1)p} \approx \frac{1}{e} \approx 37\%$$

Случайные графы [Erdős, Rényi, 1959-1961]

$$n = 50, p = 0.08$$



Эпидемиология!

Экономическая модель дружбы, [M. Jackson, 2010]

1. National Longitudinal Survey of Adolescent Health (1994 – ...): 90 000 high school students из США
2. Индивиды имеют различные типы:
 - возраст
 - раса
 - пол
 - вероисповедание
 - профессия



| Percent of Friends by Ethnicity: | Ethnicity of Students | | | |
|-------------------------------------|-----------------------|--------------|--------------|-------------|
| | White | Black | Hispanic | Others |
| | n=131 %=51 | n=96 %=38 | n= 13 %=5 | n=15 %=6 |
| White | 85 | 7 | 47 | 74 |
| Black | 4 | 85 | 46 | 11 |
| Hispanic | 4 | 6 | 2 | 4 |
| Others | 7 | 2 | 5 | 11 |

Огромное смещение в сторону своего типа для типов «white» и «black»

→ гомофилия (homophily)

Экономическая модель дружбы [M. Jackson, 2010]

1. Большие группы формируют больше внутригрупповых связей и меньше межгрупповых связей [кол-во дуг между вершинами одного/различного цвета]
2. Большие группы формируют больше связей на одного члена группы (8 против 6) [кол-во дуг, входящих в вершину]

Модель Барабаши — Альберт, 1999

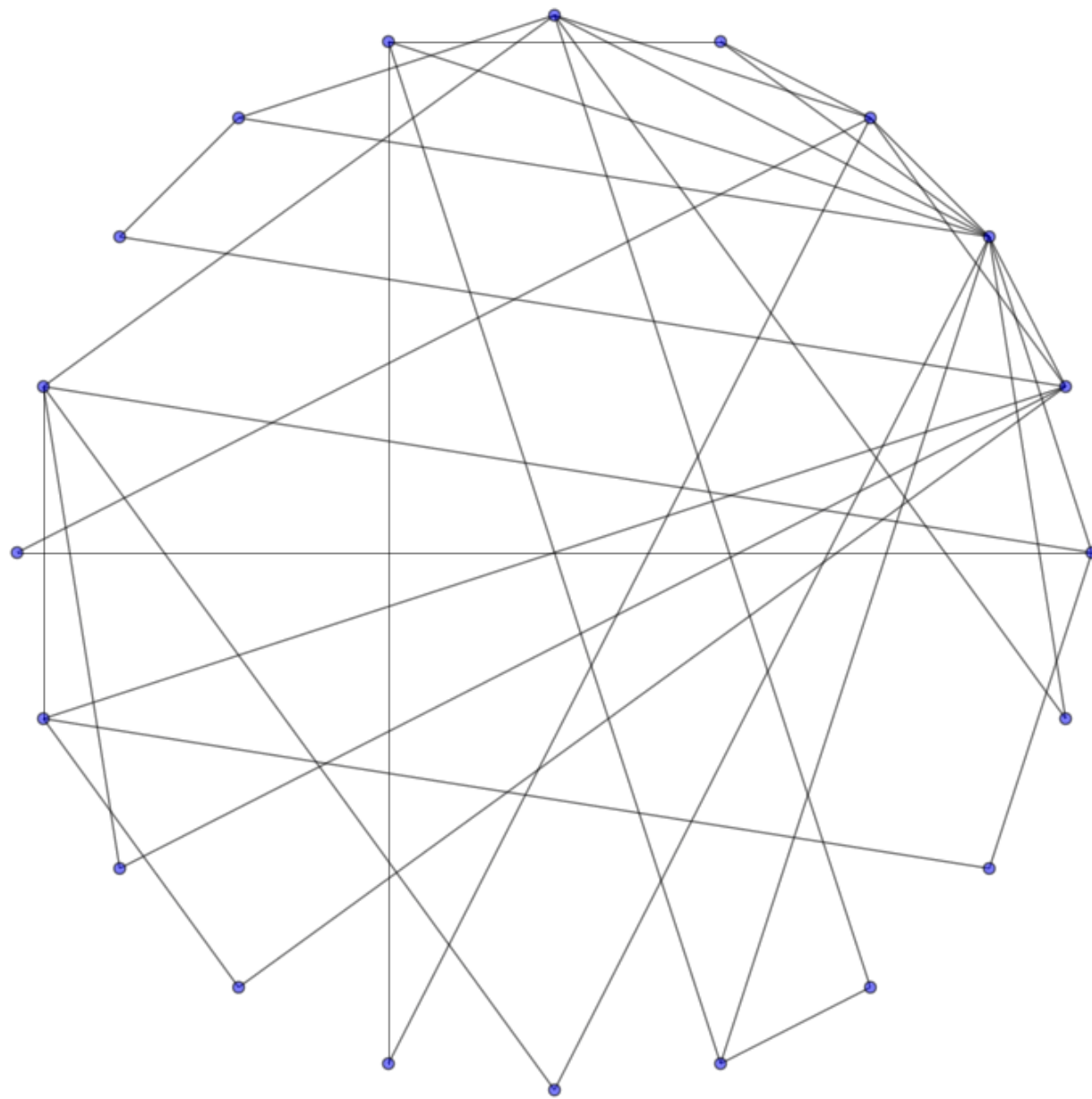
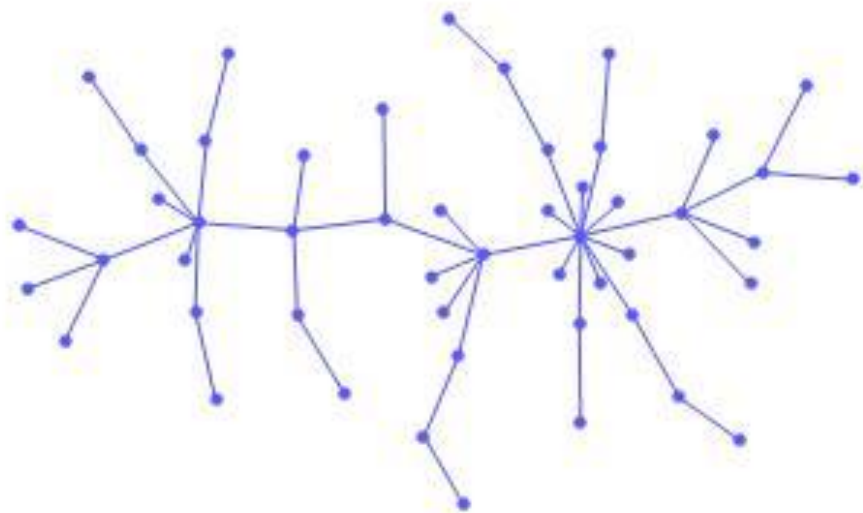
Безмасштабная сеть:

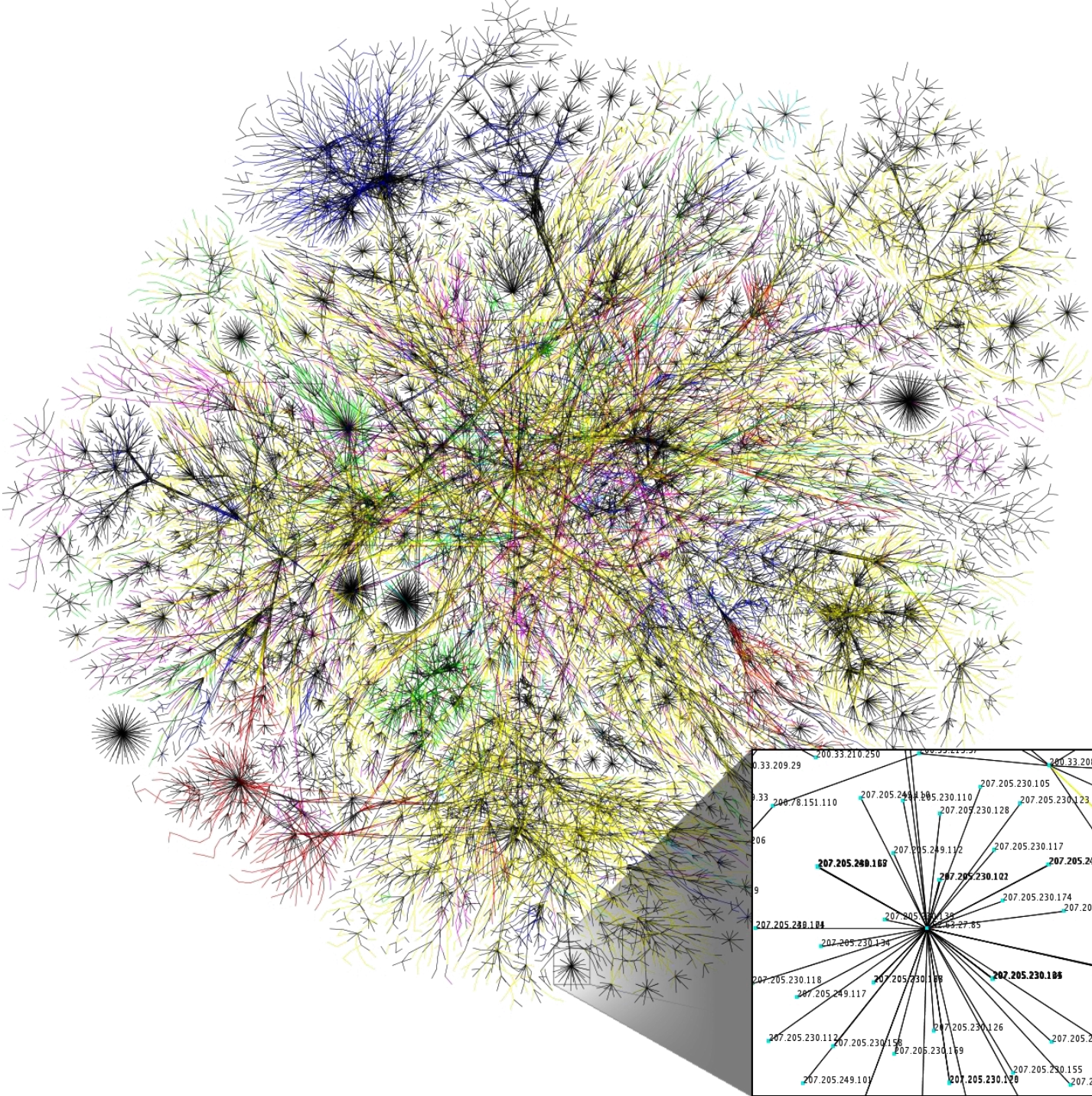
- пусть сеть начинается с $m_0 \geq 2$ вершин, степень каждой не меньше 1;
- каждый момент времени в сеть добавляется новая вершина, соединяется с i -ой существующей вершиной с вероятностью

$$p_i = \frac{k_i}{\sum k_j}$$

где k_i — степень i -ой вершины.

⇒ новые вершины имеют «предпочтение» соединяться с наиболее центральными вершинами.





Литература

1. Banerjee A., Duflo E., Jackson M. The Diffusion of Microfinance // *Science*, Vol. 341. – 2013.
2. Jackson M. *Social and Economic Networks*. – Princeton University Press. – 2010.
3. Jackson M., Currarini S., Pin P. Identifying the roles of race-based choice and chance in high school friendship network formation // *Proceedings of the National Academy of Sciences*. 107(11): 4857-4861, March 16, 2010.
4. Padgett J., Ansell C. Robust Action and the Rise of the Medici, 1400-1434 // *American Journal of Sociology*, Vol. 98, No. 6, pp. 1259-1319. – 1993.