

# Анализ социальных сетей (SNA)

Май, 2015

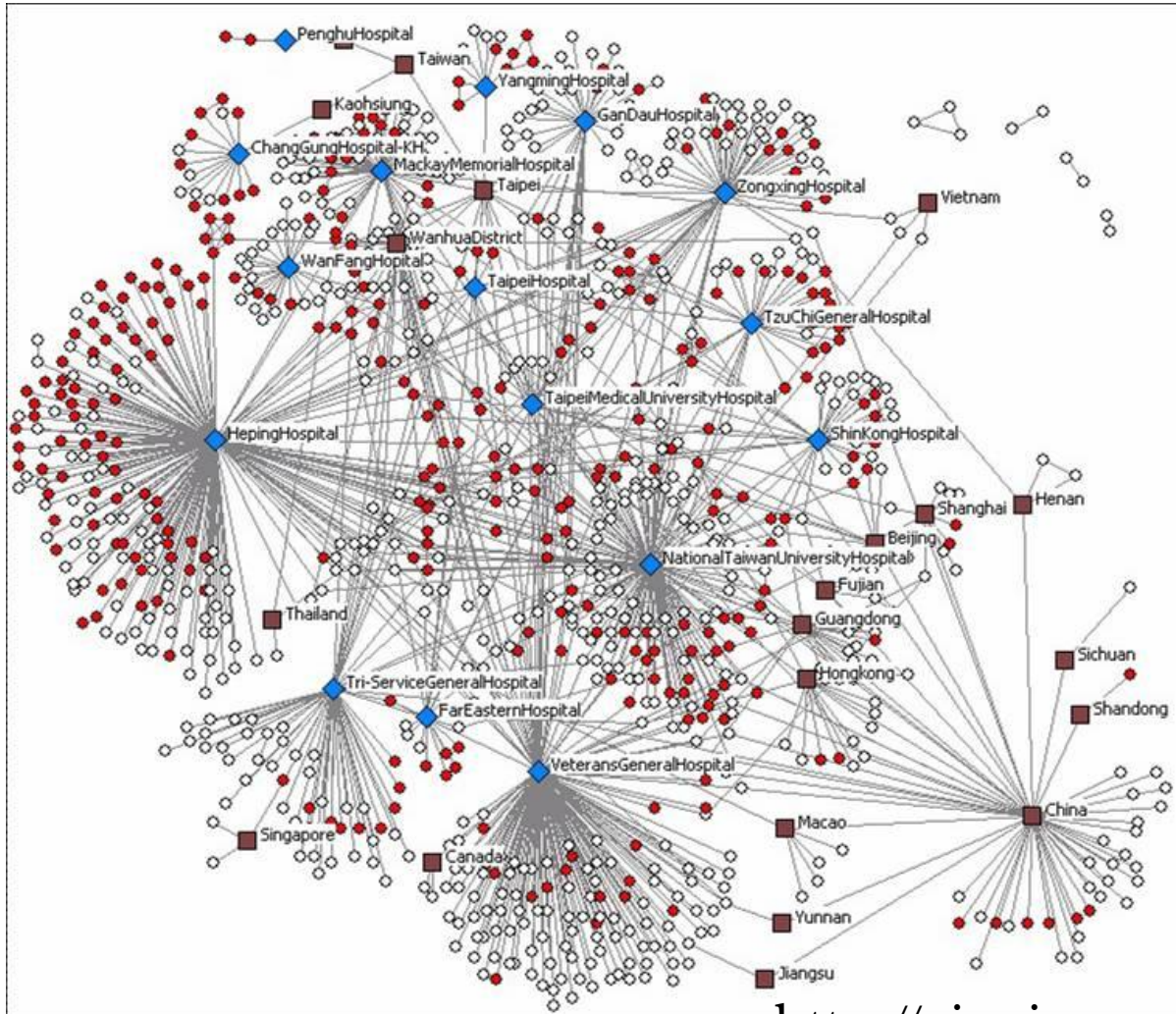
<http://prezentacija.biz>

/

# Зачем

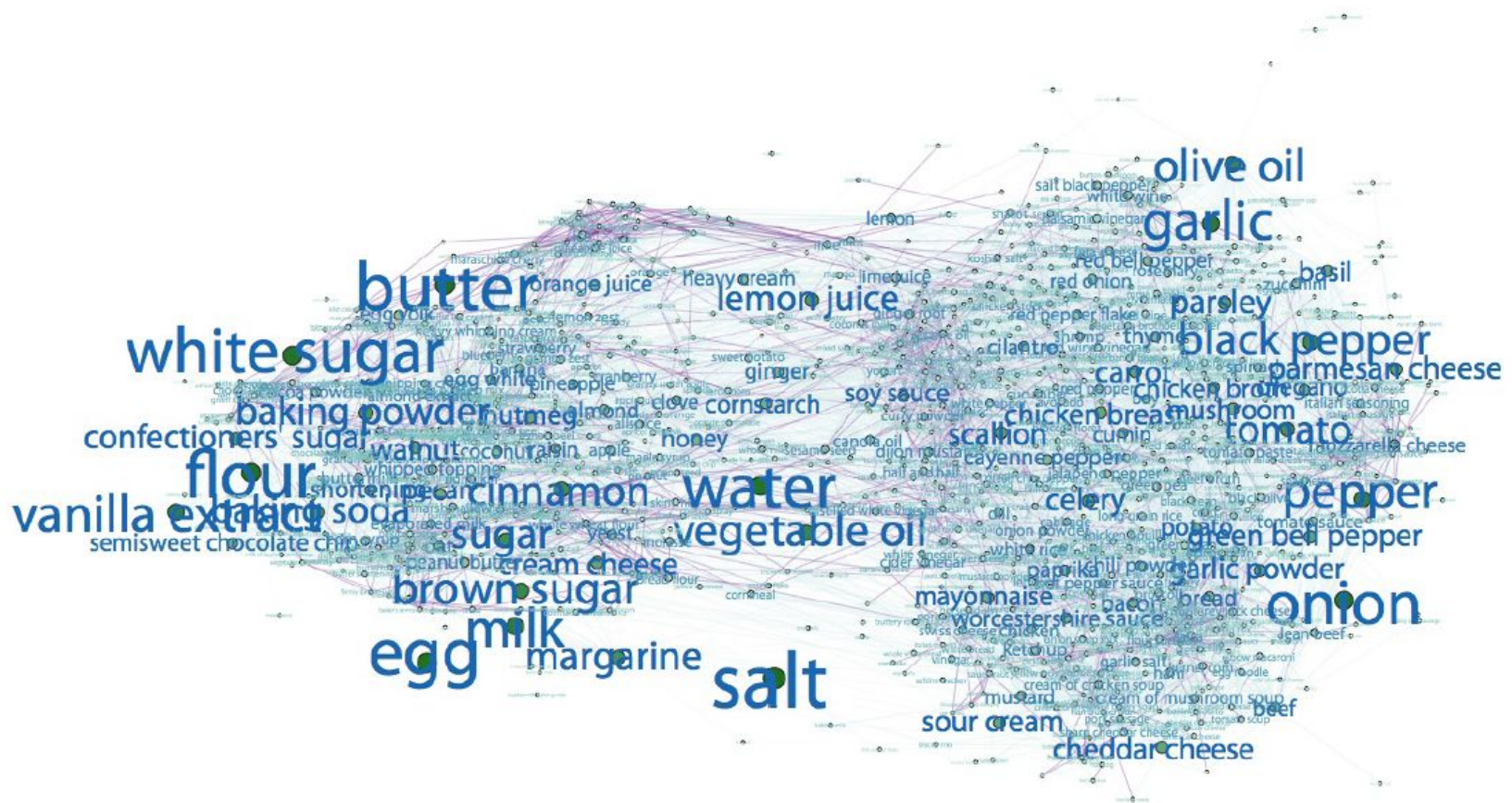
- «Все связано»
  - Изучение связей
  - Выявление влиятельных объектов
  - Распространение идей, заболеваний, информации...
- Применение в различных областях
  - социология, экономика, психология, маркетинг, история, политика, здравоохранение, литература, биология...

# Атипичная пневмония



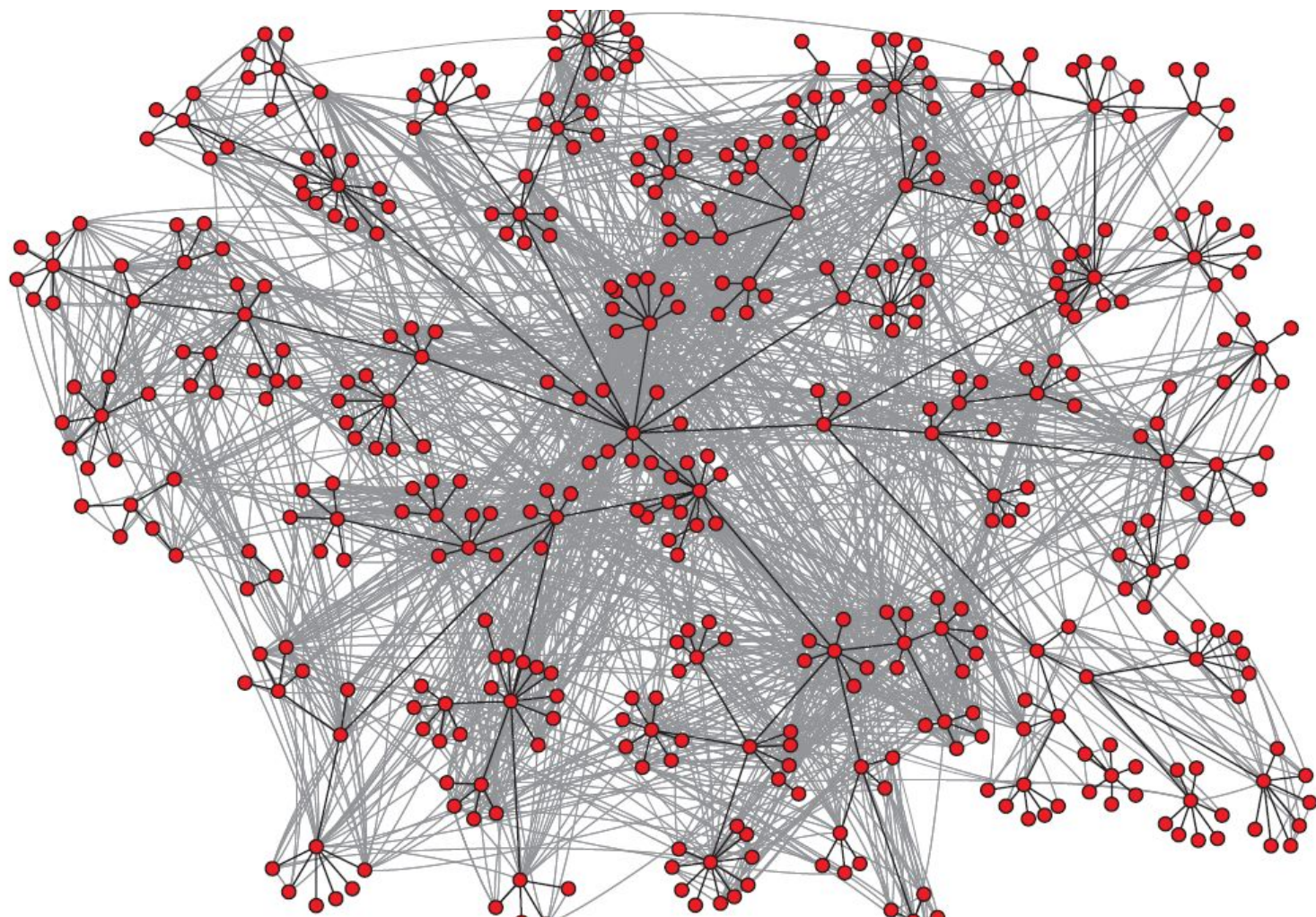
<http://ai.arizona.edu/research/bioportal/>

# Рецепты и ингредиенты



Teng C.Y. et al. Recipe recommendation using ingredient networks

# Письма в организации

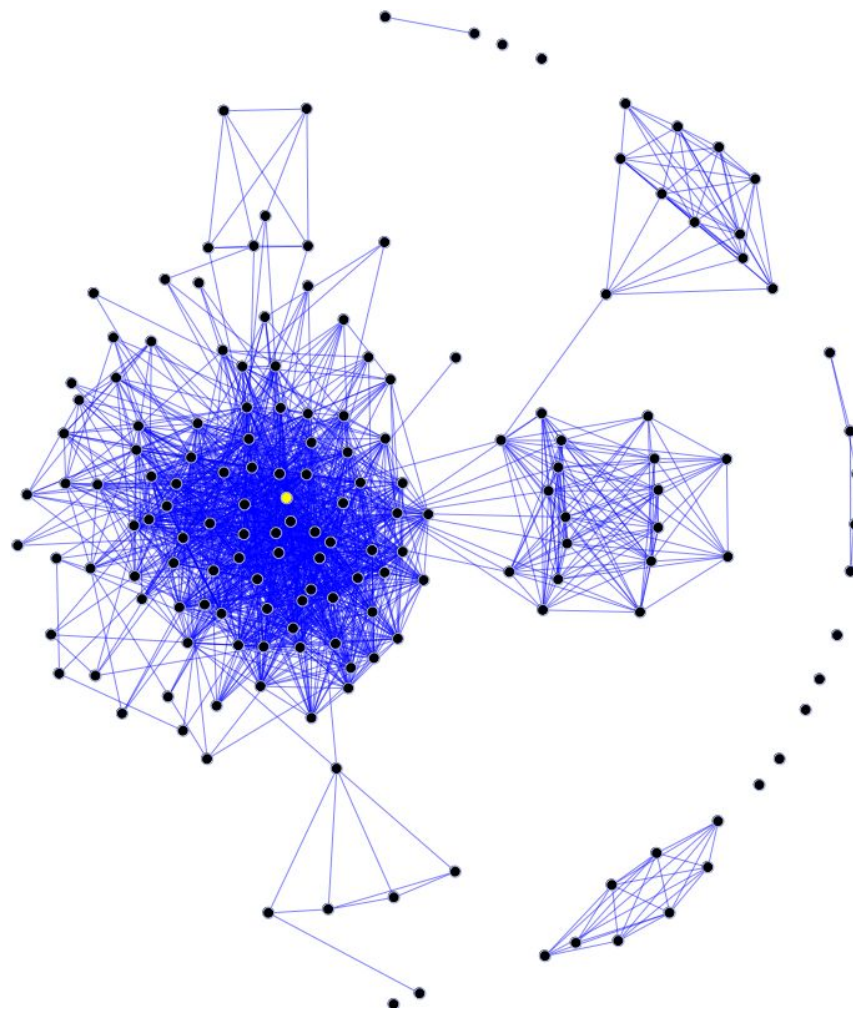
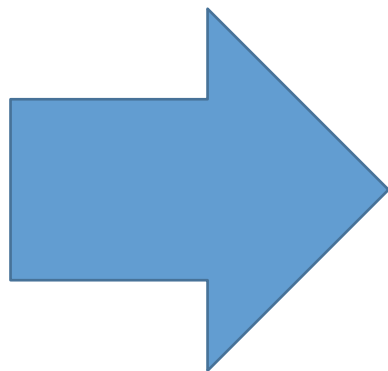


# Визуализация

---

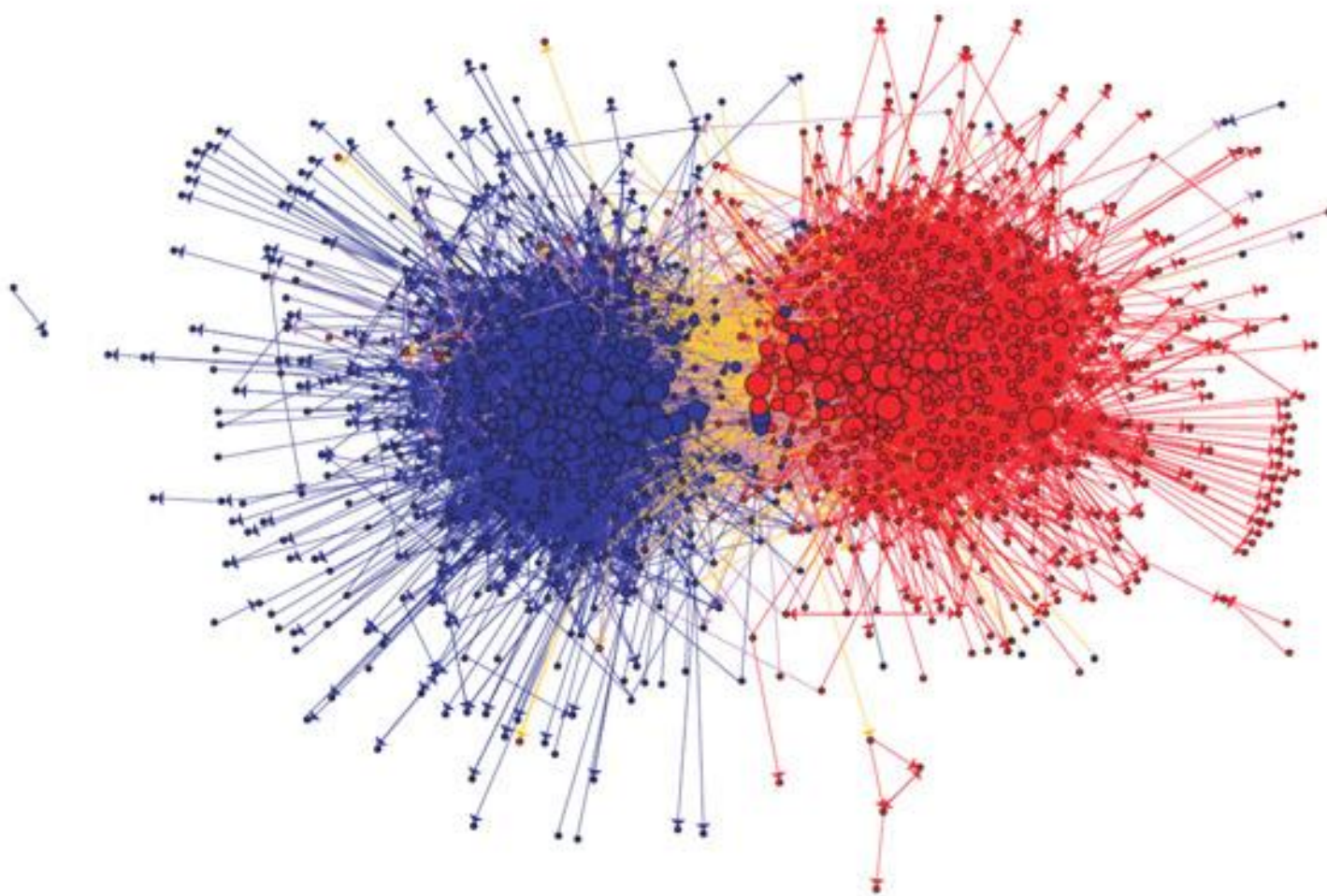
# Упрощение восприятия

Ребра	Вершины
1	I, II
2	I, III
3	II, IV
4	I, V
5	II, VI
6	III, IV
7	III, V
8	IV, VI
9	V, VII
10	VI, VII



Ребра	Вершины
1	I, II
2	I, III
3	II, IV
4	III, V
5	III, VI
6	III, VII
7	VII, VII

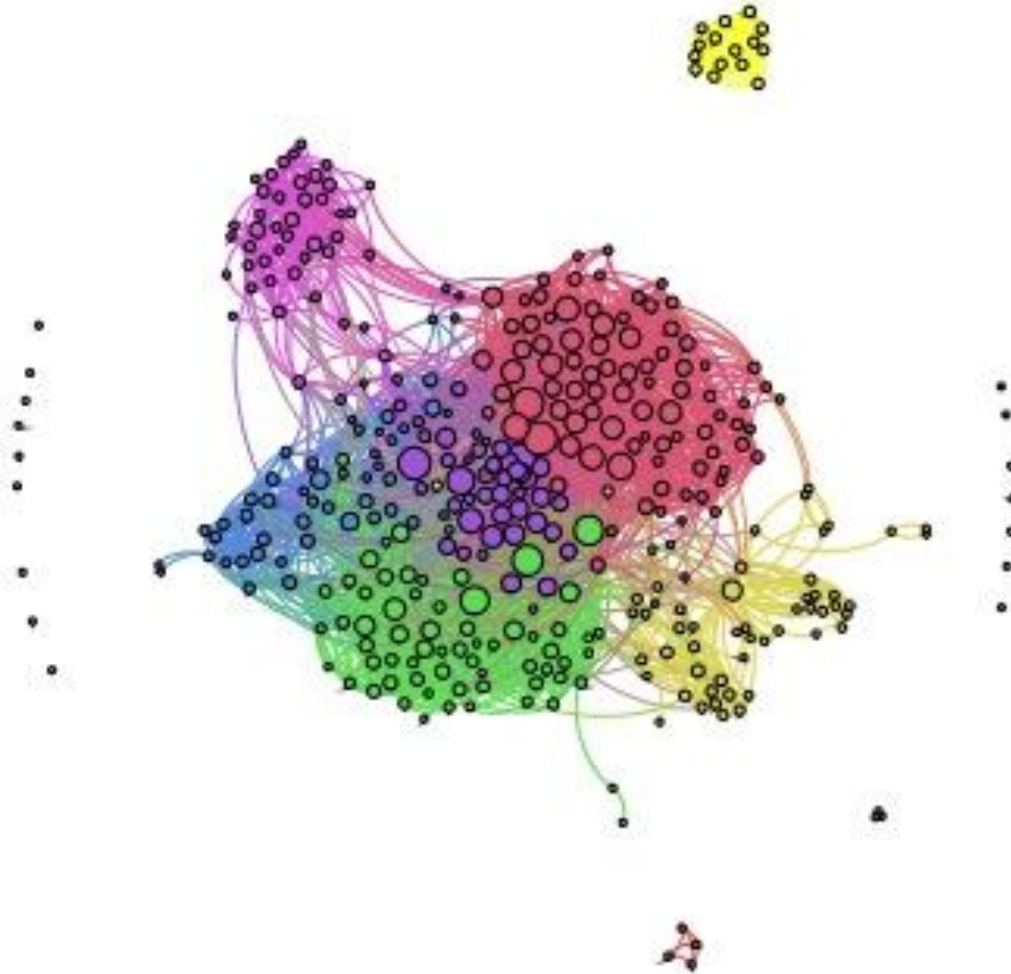
# Пример: Политические блоги



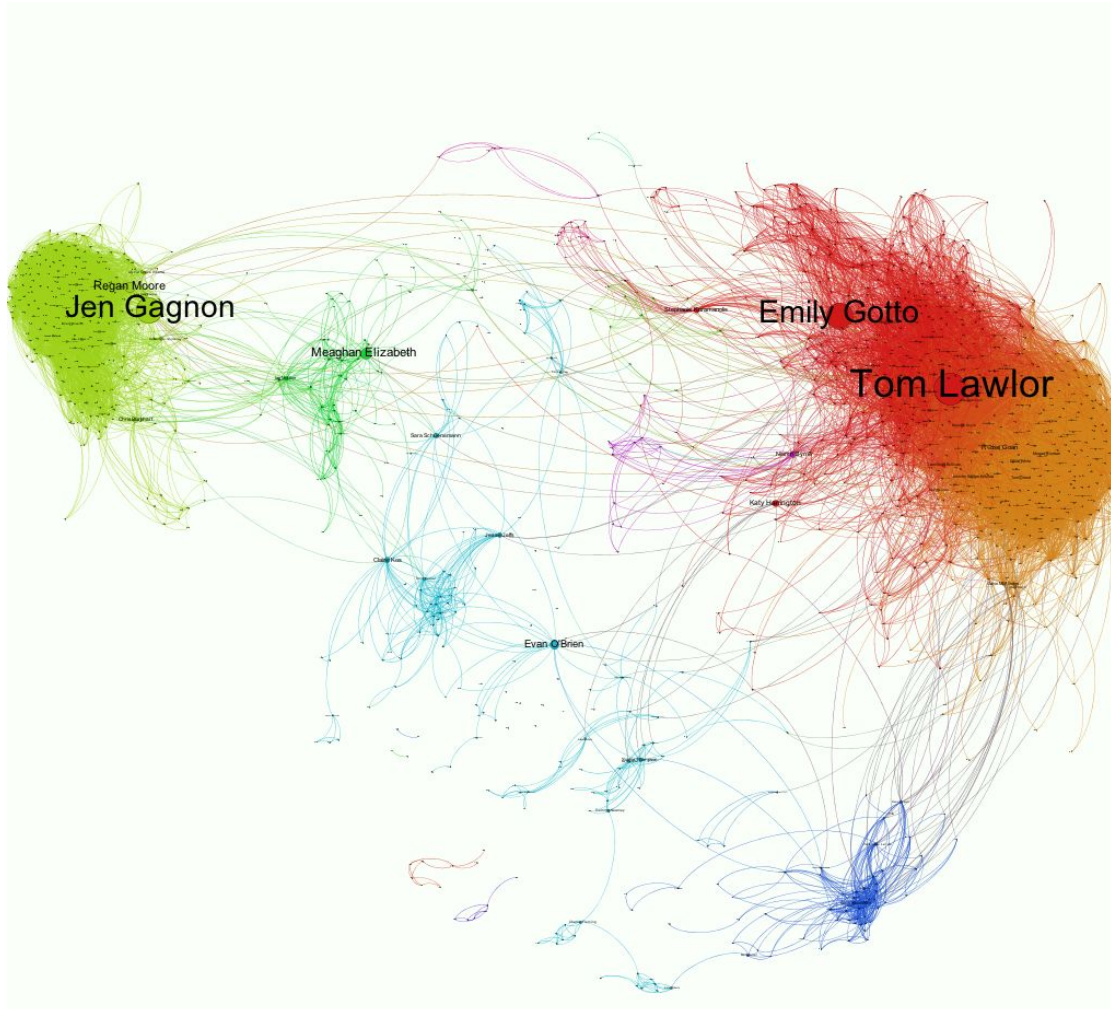


# Пример: Дружеские связи

---



# Пример: Профессиональные СВЯЗИ



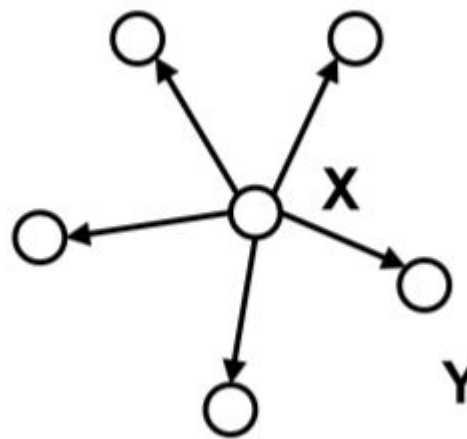
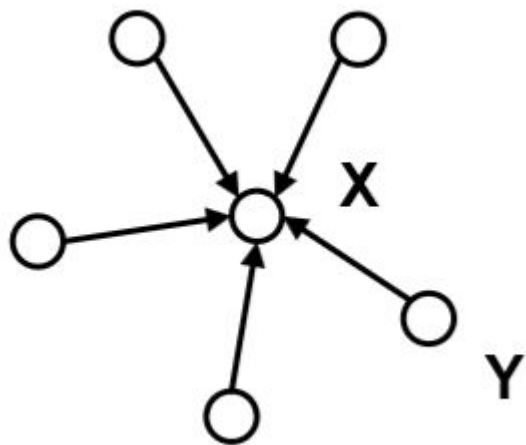
# Метрики

---

Мощность, меры центральности

# Мощность

- Входящая
- Исходящая
- Компоненты связности, главный компонент



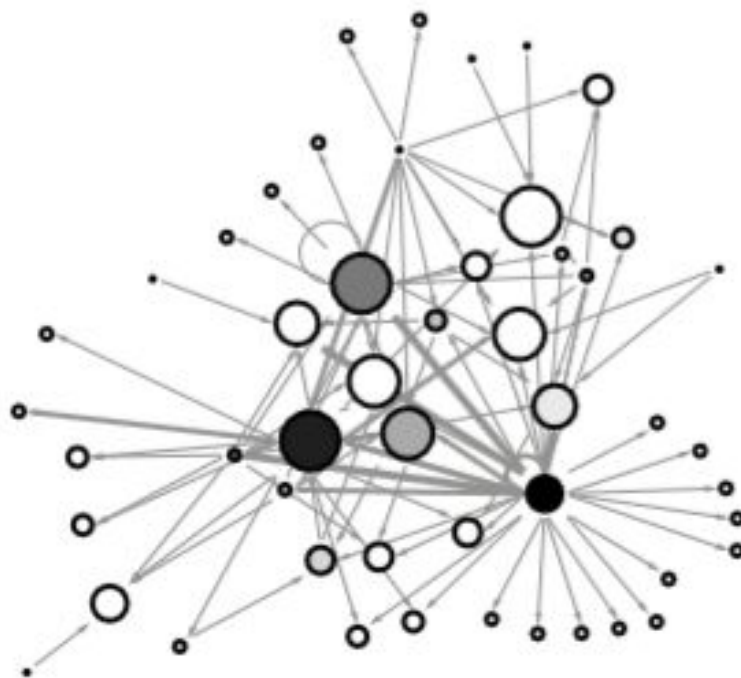
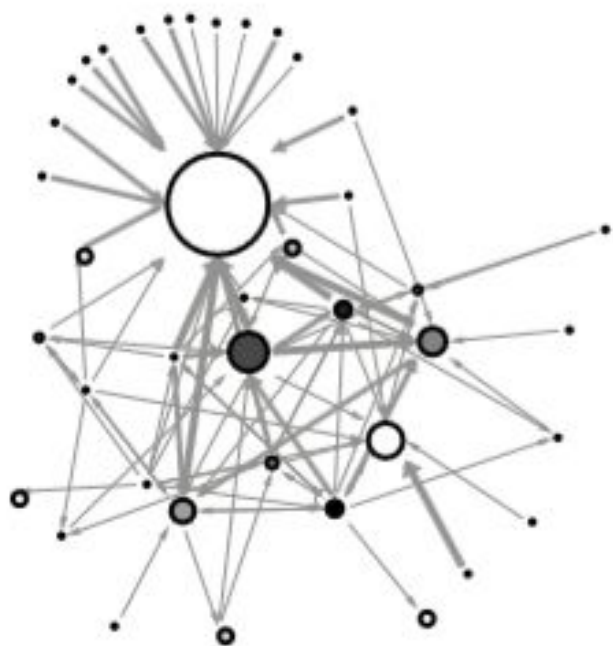
# Центральность

- есть ли выделяющиеся узлы

$$C_D = \frac{\sum_{i=1}^g [C_D(n^*) - C_D(i)]}{[(N-1)(N-2)]}$$

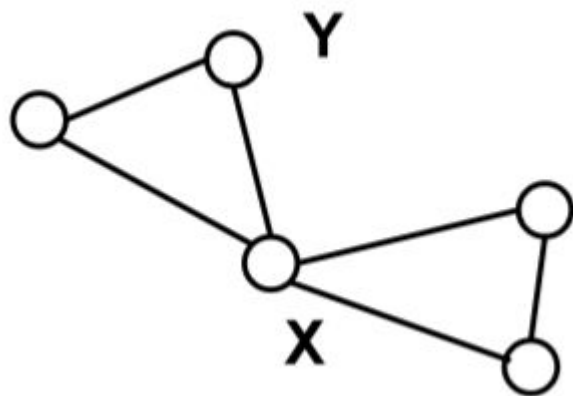
maximum value in the network

# Центральность: пример



# Мосты

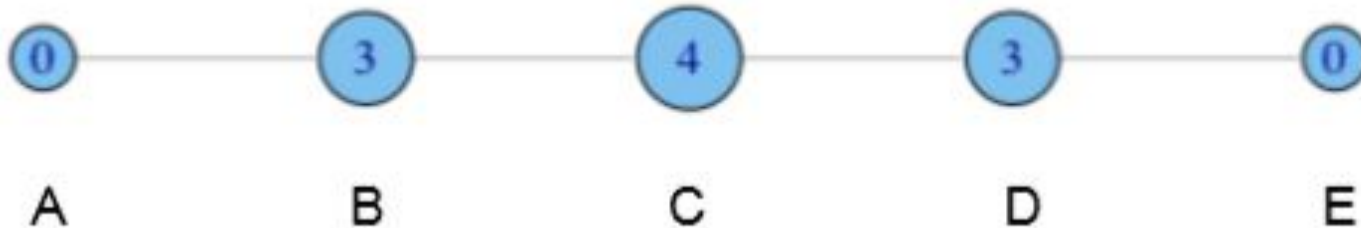
- существуют узлы с низкой мощностью, но «в выгодном месте»



# Betweenness

- число путей, проходящих через узел

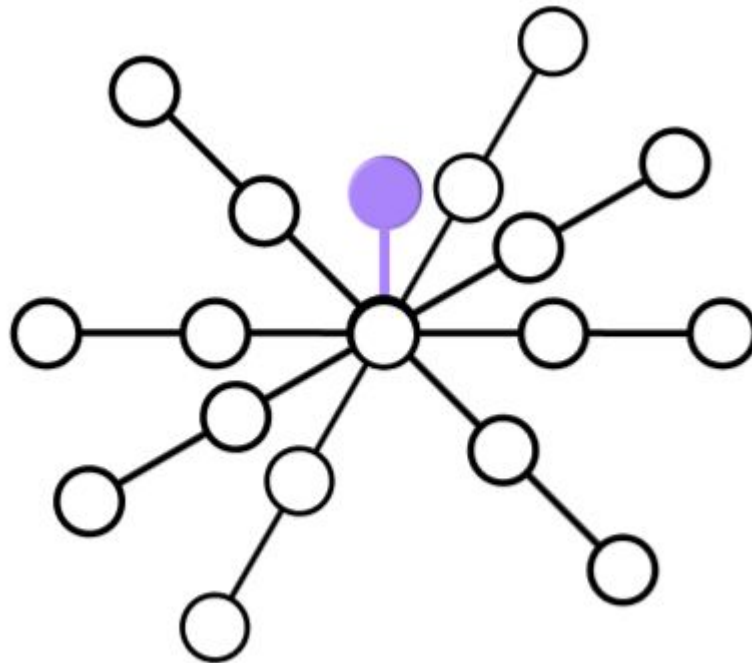
$$C_B(i) = \sum_{j < k} g_{jk}(i) / g_{jk}$$





# Близость

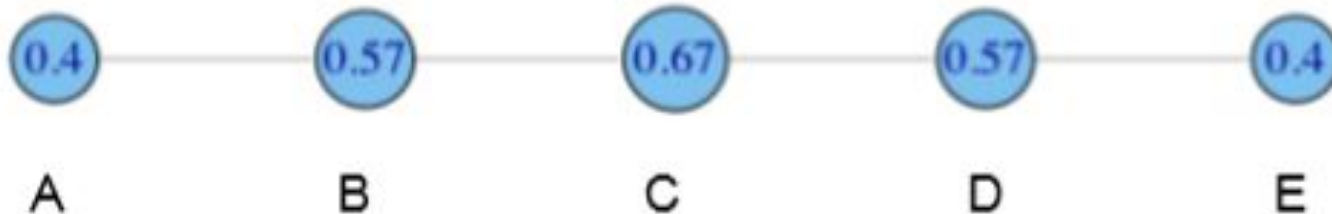
- важно быть в активном окружении



# Близость

- основана на длине кратчайшего пути между рассматриваемой вершиной и всеми остальными

$$C_c(i) = \left[ \sum_{j=1}^N d(i, j) \right]^{-1}$$



# Другие метрики

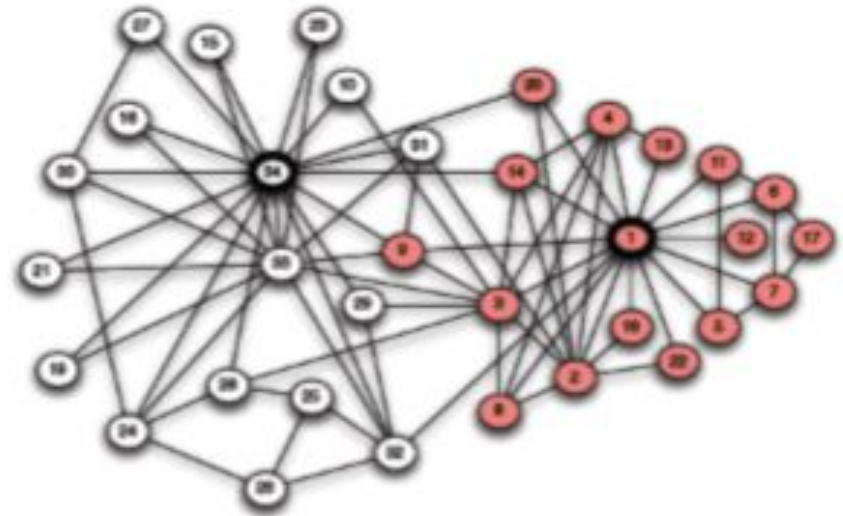
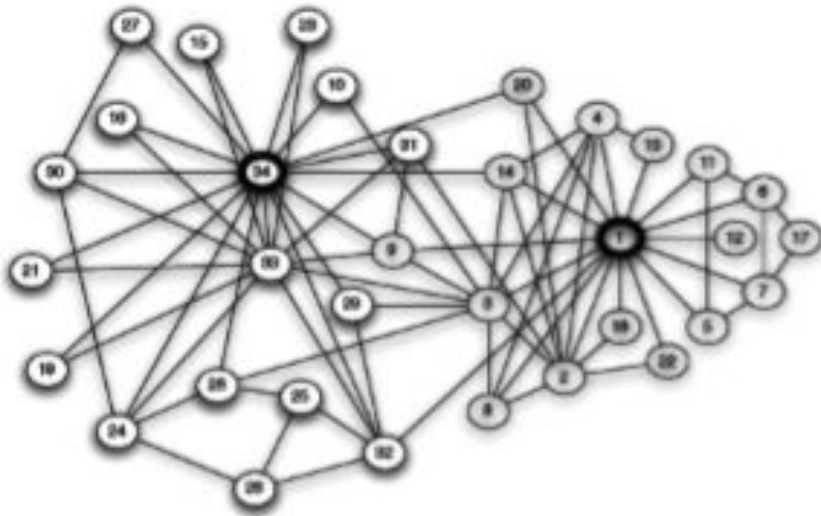
- eigenvectors
- расширение на ориентированные графы
- Page Rank
- ...

# Сообщества

---

# Выделение сообществ

- изучение структуры
- распространение информации
- формирование мнений



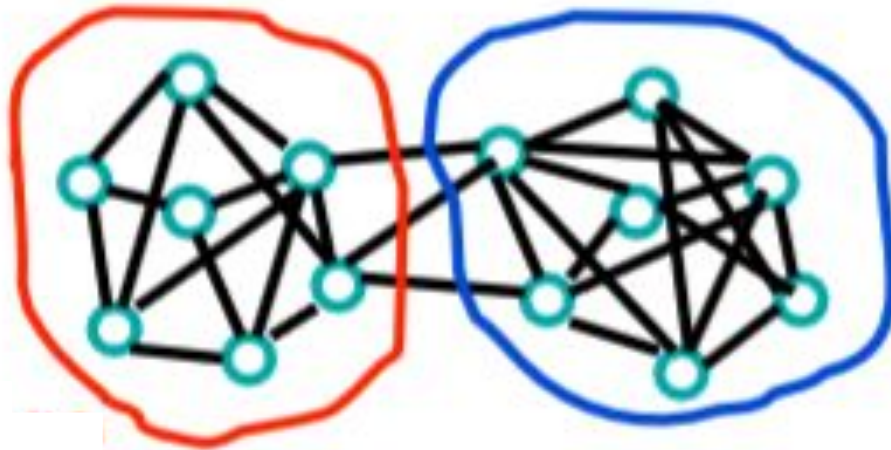
# Клики

- все узлы в клике связаны друг с другом



# к-ядра

- каждый узел связан с  $k$  других узлов



# Другие способы

- $n$ -клики: максимальное расстояние
- $r$ -клики: частота связи в группе
  
- кластеризация
- модулярность



# Случайные графы

---

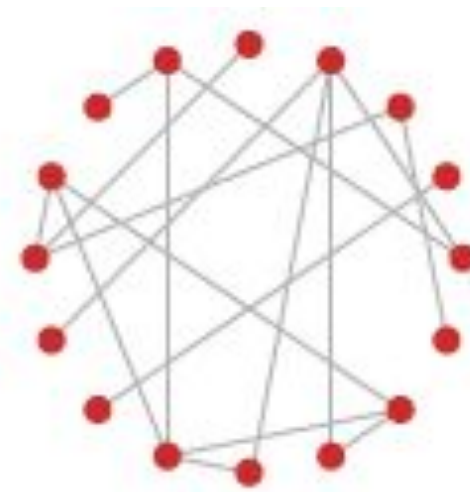
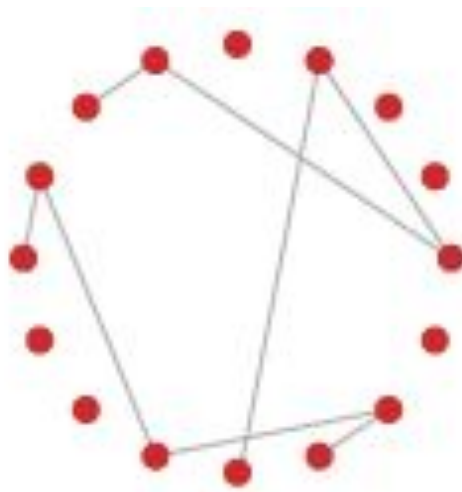
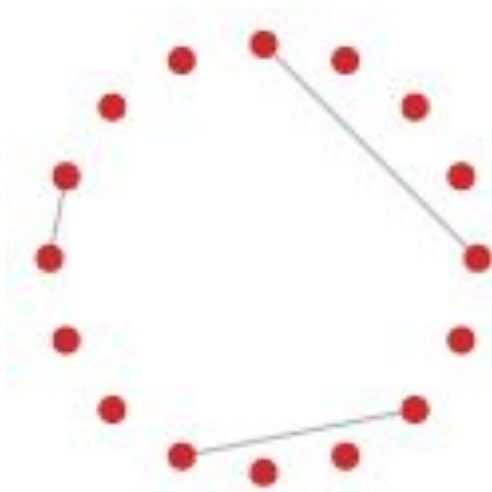
# Зачем нужна модель

- Упрощение представления
- Математические выводы
- Предсказания
- Сравнения и различия

# Граф Эрдеша-Реньи

- Неориентированный граф
- Узлы связаны случайно
- Параметры
  - $N$  – число вершин
  - $p$  – вероятность ребра
  - $M$  – общее число ребер

# Построение случайного графа



# Безмасштабные сети

- Степенное распределение по степени узла
- Соответствие сетям реального мира
- Число узлов увеличивается со временем
  - WWW
  - социальные сети
  - дружеские предпочтения
  - ...

# Модель Барабаши-Алберт

- рост сети
- предпочтительное соединение
  - чем выше степень узла, тем вероятнее, что новый узел будет присоединен к нему

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

# Построение модели



«Barabasi Albert model» участника Horváth Árpád -  
Created by the NetworkX module of the Python.

# Распространение информации

---



# Моделирование в сетях

- распространение
  - информации
  - идей
  - заболеваний
  - знаний
- координация
- мобилизация
- математические модели
- граничные значения

# Программные инструменты

- Gephi
- NetLogo
- iGraph
- Pajek
- UCINET
- NodeXL
- NetworkX (Python)
- sna (R), и другие пакеты

# Что посмотреть

- Social Network Analysis  
<https://www.coursera.org/course/sna>
- Данные для Gephi  
<https://github.com/gephi/gephi/wiki/Datasets>
- Newman M. E. J. Modularity and community structure in networks  
<http://www.pnas.org/content/103/23/8577.full>
- Freeman L. 2004. The Development of Social Network Analysis: A Study in the Sociology of Science. Empirical Press of Vancouver, BC, Canada
- Wellman B. Marin A. 2011. Social Network Analysis: An Introduction, in P. Carrington, J. Scott, eds., Handbook of Social Network. Thousand Oaks, CA: Sage.
- Wasserman S., Faust K. 2005. Social networks analysis: Methods and applications. New York: