The Small World Phenomenon:

An Algorithmic Perspective

Докладчик: Славнов Константин
МГУ ВМК ММП 3 курс

Сделано на основе презентации Брафорда Грининга

Эксперимент Милгрэма (1967)

- Выбираем получателя
- Просим случайно выбранных людей доставить письмо до цели
  - О цели известно его имя, адрес и некоторая личная информация
  - Участники эксперимента могут передавать письмо только одному человеку, с которым они хорошо знакомы
  - Цель: Доставить письмо до цели как можно быстрее

Эксперимент Милгрэма (1967)

- Результаты эксперимента:
  - 300 цепочек писем
  - 64 из них достигли цели
  - Медиана тех цепочек, которые дошли до цели примерно равна 6

Эксперимент Милгрэма (1967)

- Выводы показали два основных компонента социальных сетей
  - Очень короткий путь между двумя случайными узлами сети
  - Люди, используя только локально доступную информацию очень хорошо находят такие пути

Другие эксперименты

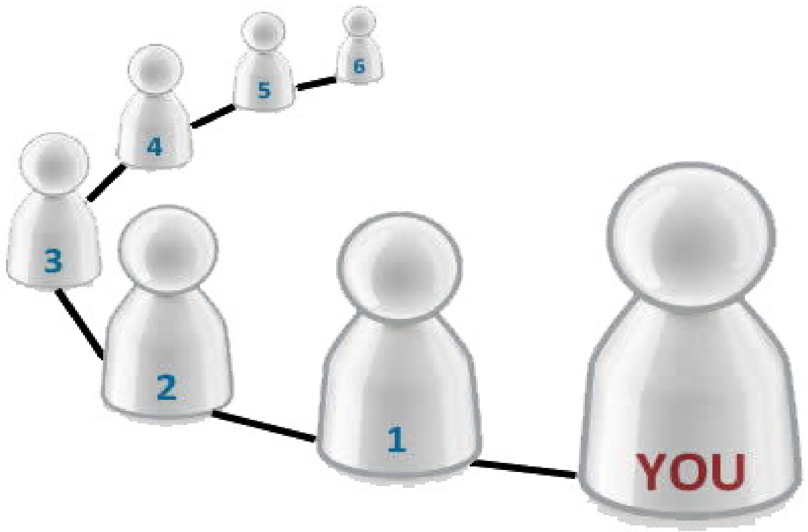
- 2001 Duncan Watts
  - e-mail сообщения. 48,000 отправителей, 19 целей. Средняя (не максимальная) длина цепочки около 6.

- 2007 Jure Leskovec и Eric Horvitz
  - Microsoft Messenger. 30 миллиардов переписок 240 миллионов людей. Средняя длинна пути у пользователей Microsoft Messenger около 6.

Что такое феномен “тесного мира”?

- Принцип, что большинство людей в обществе связаны друг с другом короткими цепочками знакомств

- Иногда называется Теория шести рукопожатий



Моделирование социальной сети

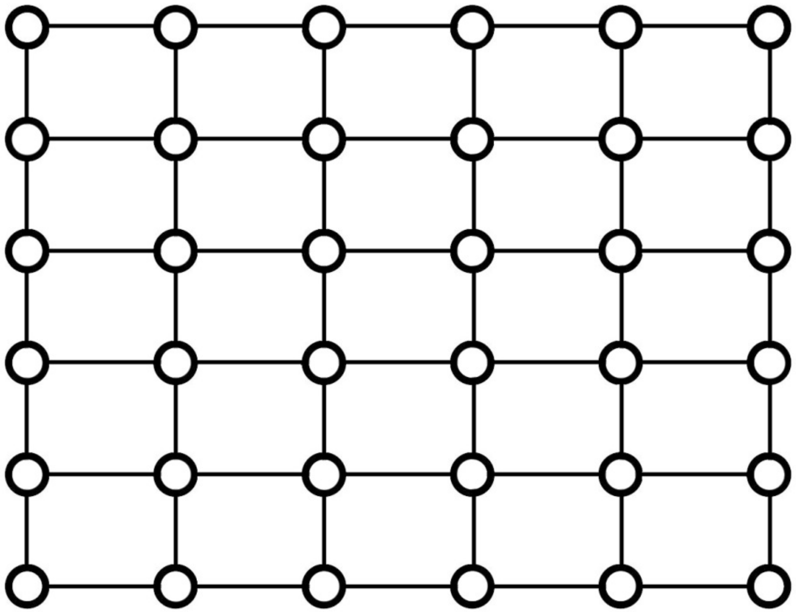
- Создание графа:
  - Узел для каждого человека
  - Если люди знают друг друга, то добавляется ребро между ними(их узлами)

- Если между почти каждой парой узлов существует “короткая” цепочка, то назовем это тесным миром

Моделирование социальной сети

- Watts – Strogatz (1998)
  - Создали модель графа для тесно связанного мира
    - Локальные связи (Local contacts)
    - Дальние связи (Long-range contacts)

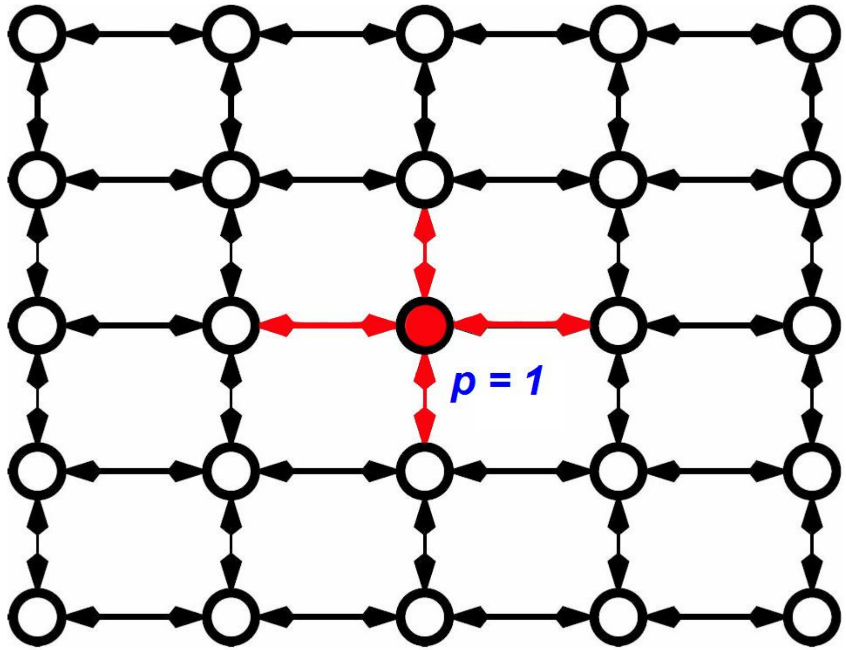
Моделирование социальной сети



- Все живут на сетке n x n
- “Решёточное расстояние” – число шагов по решетке между двумя точками
- Константы p, q

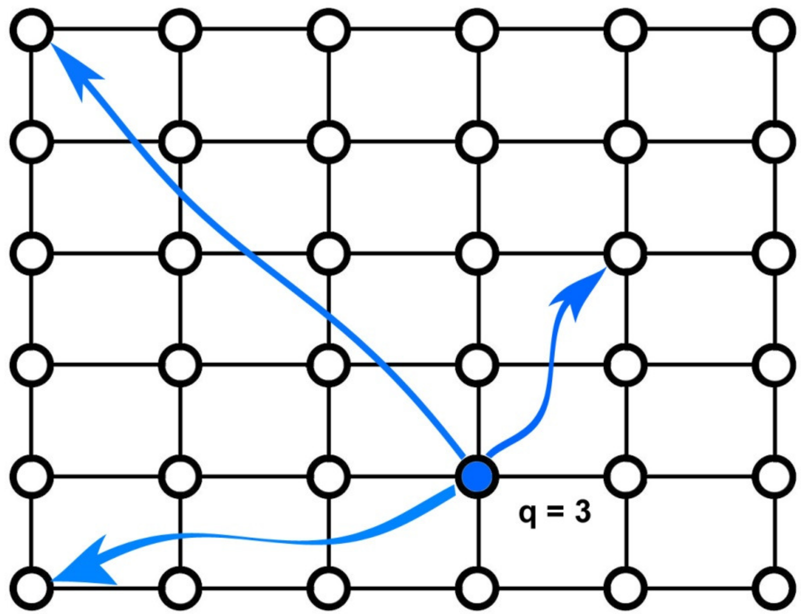
Моделирование социальной сети

- *p* - диапазон локальных связей
    - Узел соединен со всеми узлами, расстояние до которых меньше чем *p*.



Моделирование социальной сети

- *q – количество дальних связей*
  - *Добавляются ребра от узла u* до *q* других узлов с помощью независимых случайных испытаний



Моделирование социальной сети

- Watts – Strogatz (1998)
  - Показано, что небольшого количества случайности в мире (даже при $q = 1$) хватает для того, чтобы он стал тесно связанным миром

Моделирование социальной сети

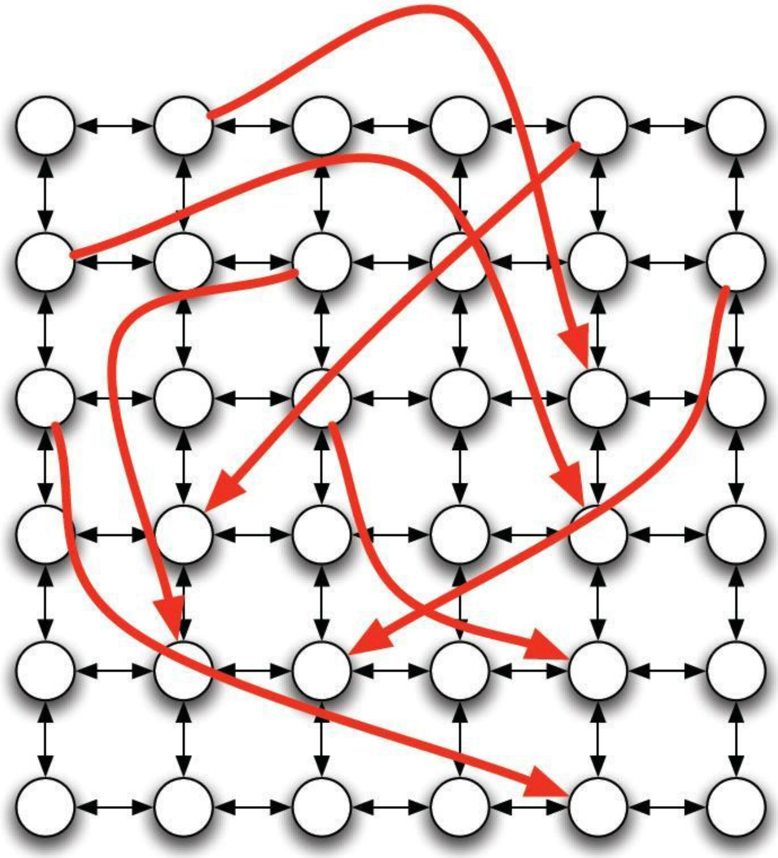
- Kleinberg (2000)
  - Почему произвольные пары незнакомых людей способны находить цепочки знакомых между друг другом используя только локальную информацию?
  - Это присуще всем тесно связанным сетям или есть свойства, которые необходимы для того, чтобы так получалось?

Моделирование социальной сети

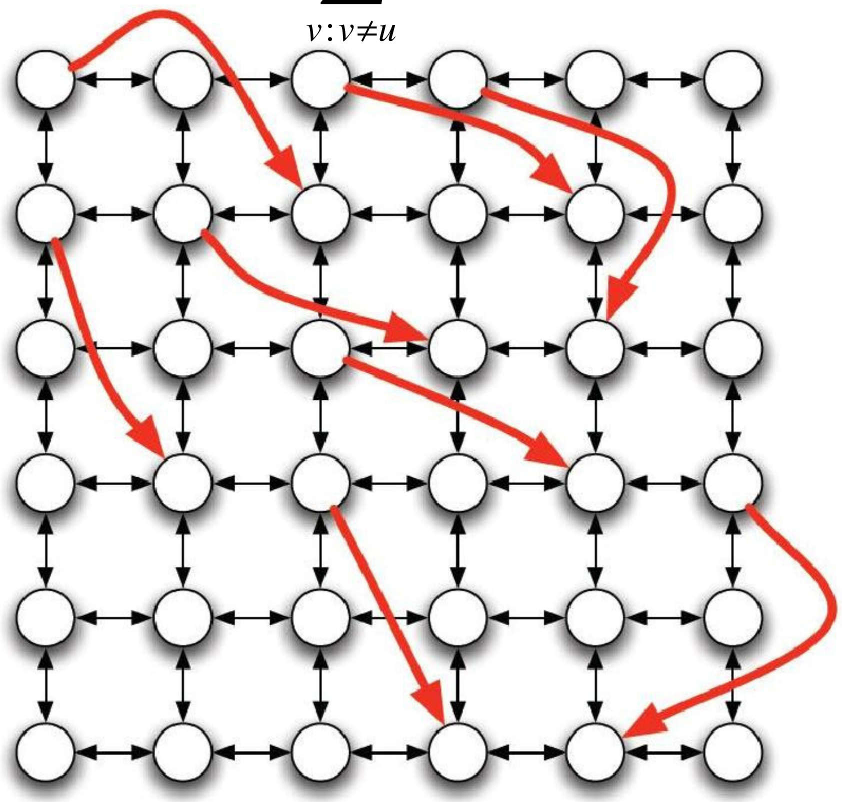
- Pr [*u* имеет дальний контакт *v*] : $\dfrac{[d(u,v)]^{-r}}{\sum_{v:v\neq u}[d(u,v)]^{-r}}$
- Бесконечное семейство сетей:
  - $r = 0$: каждая дальняя связь выбирается независимо от расстояния от u
  - С ростом r дальние связи узла u становятся более сконцентрированными в окрестности u

Моделирование социальной сети

- Pr [u имеет дальний контакт v] : $\dfrac{[d(u,v)]^{-r}}{\sum_{v:v\neq u}[d(u,v)]^{-r}}$



Маленькое r



Большое r

Алгоритмическая сторона

- Вход:
  - Граф G = (V,E)
  - Произвольные узлы *s*, *t*

- Цель:
  - Доставить сообщение от *s* до *t* коротким путем используя только локальную информацию

Алгоритмическая сторона

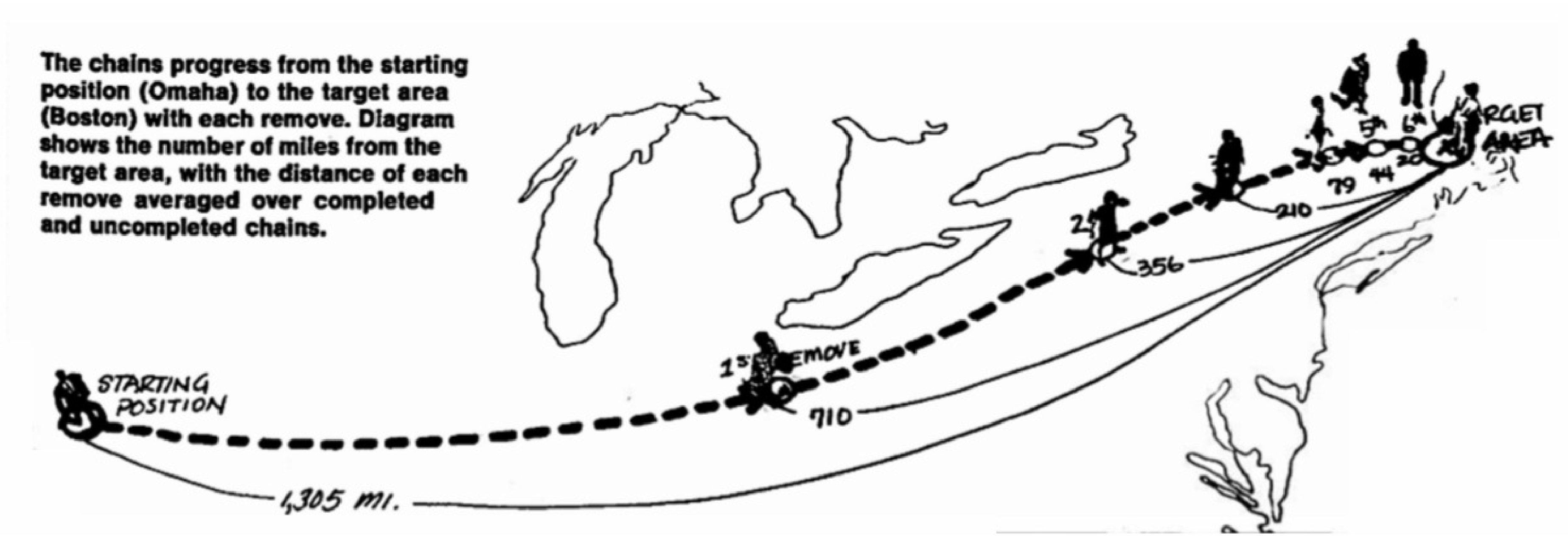
- Предположения:
  - На каждом шагу держатель сообщения *u* знает
    - Радиус локальных связей всех узлов
    - Расположение цели *t* на решетке
    - Расположение и дальние связи всех узлов, которые уже передавали сообщение
  - *u* не знает
    - Дальние связи всех узлов, которых не коснулось сообщение

$$r = 2$$

Алгоритм

- На каждом шагу текущий держатель сообщения передает его как можно ближе к цели используя свои связи

Анализ

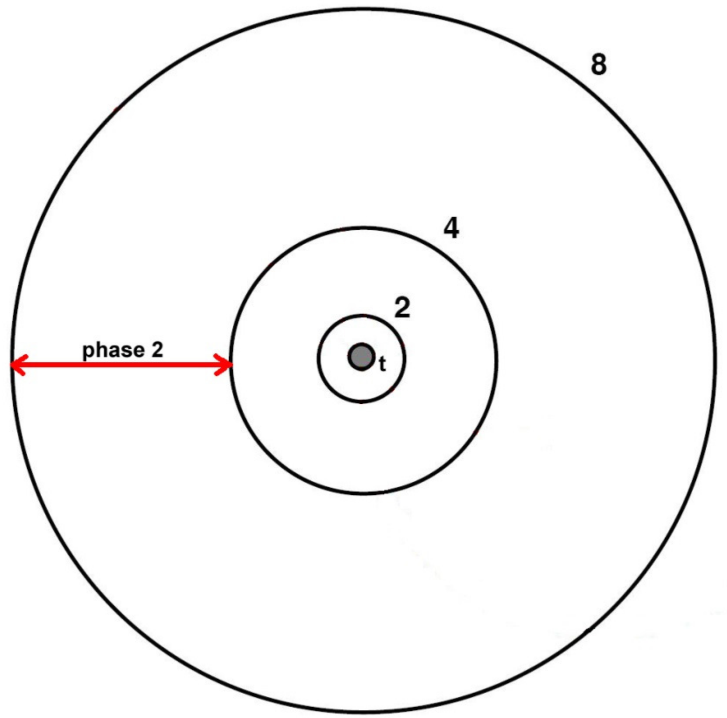


Одна из удачных цепочек в эксперименте Милгрэма

Анализ

- Алгоритм в фазе *j*:
  - На данном шаге,

    $2^j < d(u,t) \le 2^{j+1}$
  - Алгоритм в нулевой фазе:
    - Сообщение не более чем в двух шагах по сетке от цели *t*.
  - $j \le \log_2 n$.



Анализ

Вопросы:

- Как много шагов понадобится алгоритму?
- Как много шагов будет потрачено в фазу *j*?
- Какова вероятность того, что на данном шаге фаза *j* подойдет к концу?
- Какова вероятность того, что у узла *u* есть дальний контакт с узлом *v* из следующей фазы?

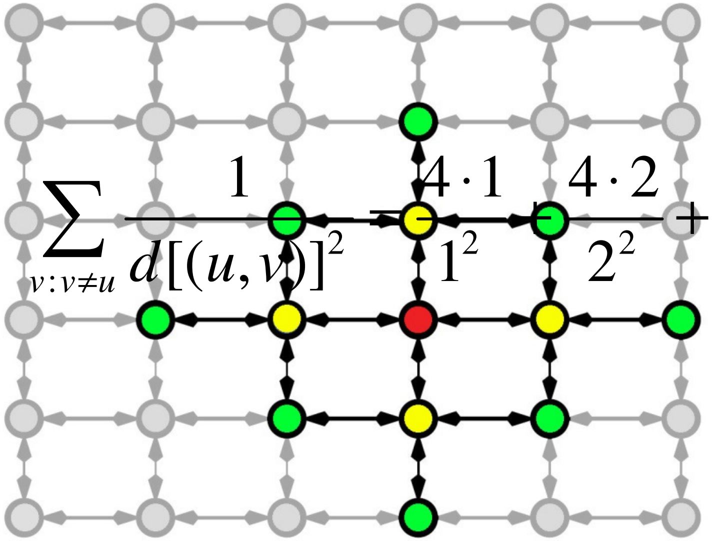
Анализ

Вопросы:

- Как много шагов понадобится алгоритму?
- Как много шагов будет потрачено в фазу j?
- Какова вероятность, что на данном шаге фаза j подойдет к концу?
- Какова вероятность того, что у u есть дальний контакт с v из *j-1* фазы?

- Pr[*u* имеет дальний контакт с *v*] ?

$$= \frac{[d(u,v)]^{-2}}{\sum [d(u,v)]^{-2}}$$



$$\sum_{v:v\neq u} \frac{1}{d[(u,v)]^2} = \frac{4\cdot 1}{1^2} + \frac{4\cdot 2}{2^2} + \frac{4\cdot 3}{3^2} + \ldots \le \sum_{j=1}^{2n-2} \frac{4j}{j^2}$$

Анализ

Вопросы:
- Как много шагов понадобится алгоритму?
- Как много шагов будет потрачено в фазу j?
- Какова вероятность, что на данном шаге фаза j подойдет к концу?
- Какова вероятность того, что у u есть дальний контакт с v из *j-1* фазы?

- Pr[*u* имеет дальний контакт с *v*] ?

$$\sum_{v:v\neq u}[d(u,v)]^{-2} \le \sum_{j=1}^{2n-2}\frac{4j}{j^2} = 4\sum_{j=1}^{2n-2}{}^{1}\!/_{j} \le 4[1+\ln(2n-2)] \le 4\ln(6n)$$

$$\ge \frac{[d(u,v)]^{-2}}{4\ln(6n)}$$

- Таким образом *u* имеет дальний контакт с *v* с вероятностью

$$\ge \frac{1}{4\ln(6n)\cdot[d(u,v)]^2}$$

Анализ

Вопросы:

- Как много шагов понадобится алгоритму?
- Как много шагов будет потрачено в фазу j?
- Какова вероятность, что на данном шаге фаза j подойдет к концу?
- Какова вероятность того, что у u есть дальний контакт с v из *j-1* фазы? $\frac{1}{4\ln(6n)\cdot[d(u,v)]^2}$

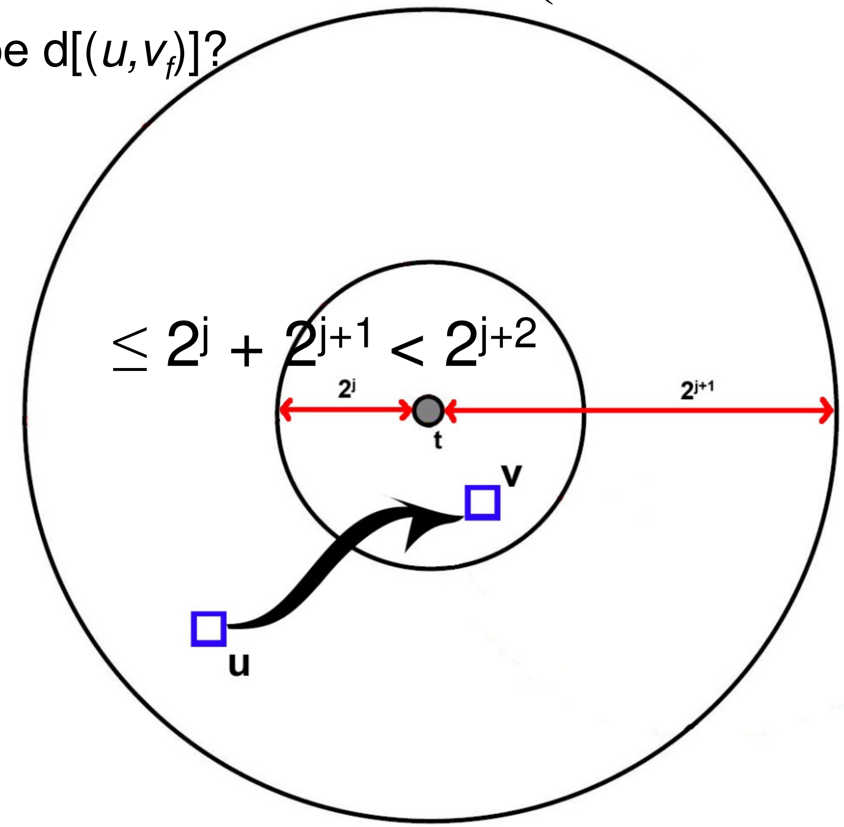
- На любом шаге, Pr[фаза *j* закончится на этом шаге] ?
  - Фаза *j* закончится этим шагом если сообщение попадет в множество B_j узлов с расстоянием не больше чем 2^j до *t*. Обозначим v_f из B_j как наиболее удаленный узел от *u*.

$$\geq |B_j| \cdot \left(\frac{1}{4\ln(6n)\cdot[d(u,v_f)]^2} \right)$$

Анализ

Вопросы:
- Как много шагов понадобится алгоритму?
- Как много шагов будет потрачено в фазу j?
- Какова вероятность, что на данном шаге фаза j подойдет к концу?
- Какова вероятность того, что у u есть дальний контакт с v из *j-1* фазы? $\frac{1}{4\ln(6n)\cdot[d(u,v)]^2}$

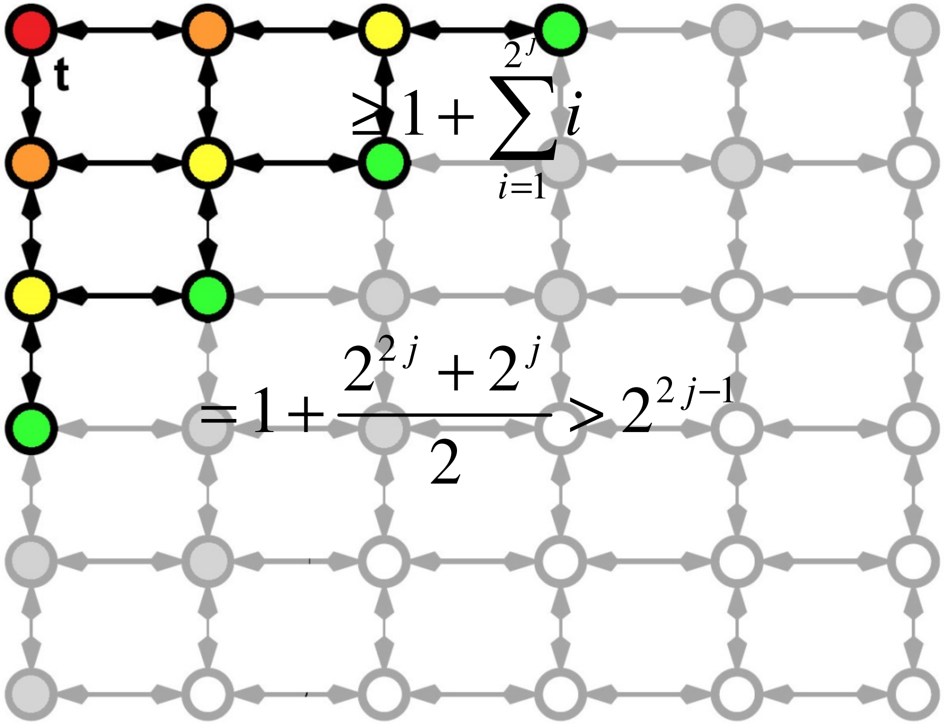
- $\Pr[\text{phase } j \text{ ends in this step}] \geq |B_j| \cdot \left(\frac{1}{4\ln(6n)\cdot[d(u,v_f)]^2}\right)$
  - Что такое d[(u,v_f)]?



Анализ

Вопросы:
- Как много шагов понадобится алгоритму?
- Как много шагов будет потрачено в фазу j?
- Какова вероятность, что на данном шаге фаза j подойдет к концу?
- Какова вероятность того, что у u есть дальний контакт с v из *j-1* фазы? $\frac{1}{4\ln(6n)\cdot[d(u,v)]^2}$

- $\Pr[\text{phase } j \text{ ends in this step}] \geq |B_j| \cdot \left(\frac{1}{4\ln(6n)\cdot 2^{2j+4}}\right)$
- Как много узлов в B_j?



$$\geq 1 + \sum_{i=1}^{2^j} i$$

$$= 1 + \frac{2^{2j} + 2^j}{2} > 2^{2j-1}$$

Анализ

Вопросы:

- Как много шагов понадобится алгоритму?
- Как много шагов будет потрачено в фазу j?
- Какова вероятность, что на данном шаге фаза j подойдет к концу?
- Какова вероятность того, что у u есть дальний контакт с v из *j-1* фазы? $\frac{1}{4\ln(6n)\cdot[d(u,v)]^2}$

- На любом шаге, Pr[фаза *j* закончится на этом шаге]?
  - Pr[*u* имеет дальний контакт с *v*] ?

$$\geq \#\ of\ nodes\ in\ B_j \cdot (probability\ u\ is\ friends\ with\ farthest\ v \in B_j)$$

$$\geq 2^{2j-1}\left(\frac{1}{4\ln(6n)\cdot 2^{2j+4}}\right) = \frac{2^{2j-1}}{4\ln(6n)\cdot 2^{2j+4}} = \frac{1}{128\ln(6n)}$$

Анализ

Вопросы:
- Как много шагов понадобится алгоритму?
- Как много шагов будет потрачено в фазу j?
- Какова вероятность, что на данном шаге фаза j подойдет к концу? $\geq \frac{1}{128 \ln(6n)}$
- Какова вероятность того, что у u есть дальний контакт с v из *j-1* фазы? $\frac{1}{4 \ln(6n) \cdot [d(u,v)]^2}$

- Как много шагов будет затрачено на фазу *j*?
  - Пусть X_j это случайная величина, которая показывает количество шагов, затраченное на фазу *j*.
    - X_j имеет геометрическое распределение с вероятностью успеха

$$\geq \frac{1}{128 \ln(6n)}$$

Анализ

Вопросы:
- Как много шагов понадобится алгоритму?
- Как много шагов будет потрачено в фазу j?
- Какова вероятность, что на данном шаге фаза j подойдет к концу? $\geq \frac{1}{128\ln(6n)}$
- Какова вероятность того, что у u есть дальний контакт с v из *j-1* фазы? $\frac{1}{4\ln(6n)\cdot[d(u,v)]^2}$

- Как много шагов будет затрачено на фазу *j*?
  - Т.к. X_j имеет геометрическое распределение

$$E[X_j] = \frac{1}{p} \leq \frac{1}{1/128\ln(6n)} = 128\ln(6n)$$

Анализ

Вопросы:

- Как много шагов понадобится алгоритму?
- Как много шагов будет потрачено в фазу j?

  $\leq 128 \ln(6n)$
- Какова вероятность, что на данном шаге фаза j подойдет к концу?

  $\geq \dfrac{1}{128 \ln(6n)}$
- Какова вероятность того, что у u есть дальний контакт с v из *j-1* фазы?

  $\geq \dfrac{1}{4 \ln(6n) \cdot [d(u,v)]^2}$

- Как много шагов требуется алгоритму?
  - Пусть X это случайная величина обозначающая количество шагов, затраченных алгоритмом.
  - Из линейности мат. ожидания получаем

$$E[X] \leq (1 + \log n)(128 \ln(6n)) = O(\log n)^2$$

Анализ

Вопросы:

- Как много шагов понадобится алгоритму?
- Как много шагов будет потрачено в фазу j?

  $\leq 128 \ln(6n)$

- Какова вероятность, что на данном шаге фаза j подойдет к концу?

  $\geq \dfrac{1}{128 \ln(6n)}$

- Какова вероятность того, что у u есть дальний контакт с v из *j-1* фазы?

  $\geq \dfrac{1}{4 \ln(6n) \cdot [d(u,v)]^2}$

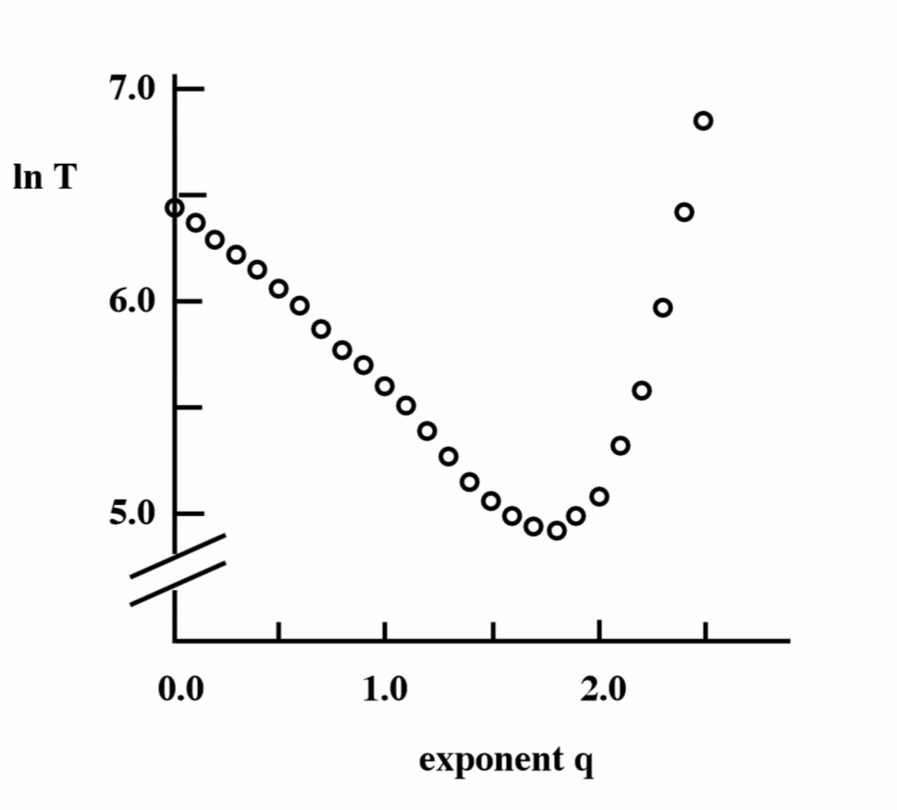
- Если r = 2, ожидаемое время доставки:

$$O(\log n)^2$$

$$r \neq 2$$

Сводка результатов

- $0 \le r < 2$: ожидаемое время доставки любого децентрализованного алгоритма поиска $\Omega(n^{(2-r)/3})$.
- $r > 2$: ожидаемое время доставки любого децентрализованного алгоритма поиска $\Omega(n^{(r-2)/(r-1)})$.



Моделирование децентрализованного поиска. Каждая точка – среднее от 1000 запусков поиска на сетке с 400 миллионами узлами

Интуитивное объяснение

- В зависимости от величины r
  - $r = 0$ нет запутанности, которая обеспечивает ускорение передачи сообщения.
  - $0 < r < 2$: недостаточная запутанность для быстрой доставки.
  - $r > 2$: с ростом r, сеть становится более локализованной.
  - $r = 2$: хорошее сочетание “географической” запутанности без слишком большой локализации.

Источники

- Easley D., Kleinberg J. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge University Press, 2010.
- Kleinberg, J. *The Small-World Phenomenon: An Algorithmic Perspective.* Proc. 32nd ACM Symposium on Theory of Computing, 2000
- Kleinberg, J. *Navigation in a Small World.* Nature 406(2000), 845.

- Оригинальная презентация: http://crab.rutgers.edu/~rajivg/studentTalks/smallworld.ppt
- http://en.wikipedia.org/wiki/Six_degrees_of_separation