

*«Мир Тесен» по Джону
Клайнбергу*

Мельников Иван
Еремин Юрий

Краткое содержание

1. Введение

2. История

3. Сетевая модель

4. Сети, поддерживающие эффективный
поиск

5. Выводы

6. Открытые вопросы

7. Ссылки

Введение

- “Мир тесен”
 - тема анекдотических исследований и фольклора
 - часто бывает, что мы встречаем незнакомца и оказывается, что у нас есть общий знакомый

Введение(2)

- задача поиска информации
 - поведение пользователей Web
 - поведение агентов
 - поисковые протоколы (Gnutella, Freenet)

Эксперимент Стэнли Милграма

- проведенный в 1960х, остается одним из самых удачных в понимании проблемы
- человек из Небраски должен был передать письмо человеку, живущему в Массачусетсе
- шесть ступеней разделения людей в США

Открытия Милграма

- короткие пути в сетях знакомств существуют
- люди могут находить эти пути, зная только информацию о конечной цели

Исследования Пула и Кочена

- случайные сети имеют маленький диаметр
- если А и Б два индивидуума с общим другом, то скорее всего они сами друзья.
- НО, очень разветвленная сеть знакомств, не имеет малого диаметра

Модель Ватса и Строгатца

- балансирует между ограничениями разветвленности сети знакомств и диаметра сети
- пример - «сетчатый круг»
- простая структура, но при этом несколько ребер произведены случайным процессом, который эта структуры не описывает

Исследования Джона Клайнберга

- Почему незнакомые люди могут найти, соединяющую их короткую цепь знакомств?
- Существуют скрытые навигационные ключи, лежащие в социальной сети
- Децентрализованные алгоритмы

Открытия Джона Клайнберга

- существующих моделей недостаточно, чтобы объяснить успех децентрализованного алгоритма
- для одной из моделей класса обобщенных сетей Ватса и Строгатца существует децентрализованный алгоритм, способный находить короткие пути с большой вероятностью.
- существует уникальная модель в этом семействе, для которой децентрализованные алгоритмы эффективны.

Другие работы по теме

- как индивидуумы выбирают следующего адресата
- Бернард и Килворф : «обратные эксперименты тесного мира»
- Вайт – «смерть» цепи
- Хантер и Шотланд - прохождение цепи по разным социальным категориям людей
- Альберта, Йонг и Барабаси - способность агентов находить пути

Сетевая модель

- Описание модели
- Децентрализованные алгоритмы
- Результаты применения модели
- k – мерная сеть

Описание модели

- квадратная сетка

$$n \times n, \{(i; j) : i \in \{1; 2; \dots; n\}; \\ j \in \{1; 2; \dots; n\}\}$$

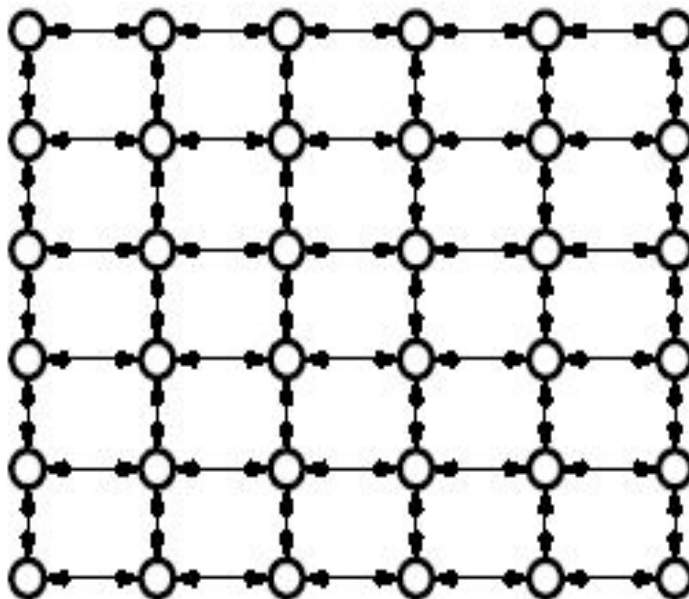
- сетчатое расстояние

$$d((i; j); (k; l)) = |k - i| + |l - j|$$

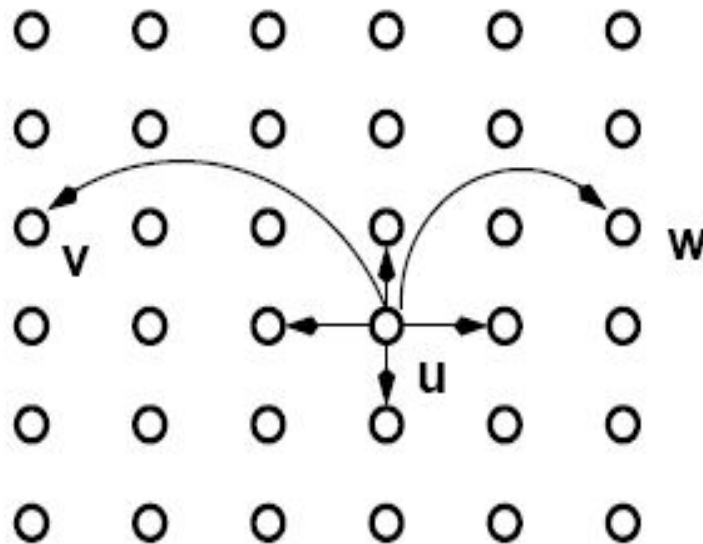
Описание модели(2)

- $p \geq 1$ - локальные контакты
- $q \geq 0$ - удаленные контакты
- $r \geq 0$
- $[d(u; v)]^{-r}$ - вероятность ребра из u в v
- $[d(u; v)]^{-r} / \sum_v [d(u; v)]^{-r}$ - введем такую величину, обратное распределение степени r

Пример: Сеточная модель



Пример: Сеточная модель (2)



Контакты узла u при $p = 1$, $q = 2$, где v и w – дальние контакты

Описание модели(3)

- считая p и q фиксированными константами получаем однопараметрическое семейство сетей, зависящее от показателя r
- r – базовый параметр, измеряющий как широко разветвлена данная сеть
- при $r = 0$ получается модель Ватса Строгаца

Децентрализованные алгоритмы

На каждом шаге держатель сообщения знает:

- множество локальных контактов
- местоположение цели на решетке
- * местоположение и дальние контакты всех узлов, которые были держателями сообщения

Децентрализованные алгоритмы

- Ожидаемое время доставки
 - ожидаемое количество шагов по пути
 - порождаем граф в соответствии с обратным распределением степени r
 - начальная и целевая точки выбираются случайно равномерно

Результаты применения модели

- **Теорема 1:**

Существует константа a_0 , зависящая от p и q , не зависящая от n , такая что при $r = 0$, ожидаемое время доставки любого децентрализованного алгоритма не меньше $a_0 n^{2/3}$

Результаты применения модели (2)

- **Теорема 2:**

Существует децентрализованный алгоритм A и константа a_2 ,
независящая от n , так что при $r = 2$ и
 $p = q = 1$, ожидаемое время доставки
не больше $a_2(\log n)^2$.

Фундаментальное следствие

- когда дальние контакты создаются процессом, связывающих их определенным образом с геометрией решетки поиск эффективен

Главные предположения теорем

- В первой теореме равномерное распределение не позволяет алгоритму использовать скрытые «ключи» геометрии решетки
- алгоритм A второй теоремы: на каждом шаге держатель сообщения выбирает контакт наиболее близкий к цели в смысле сетчатой метрики

Теорема

- $0 \leq r < 2$

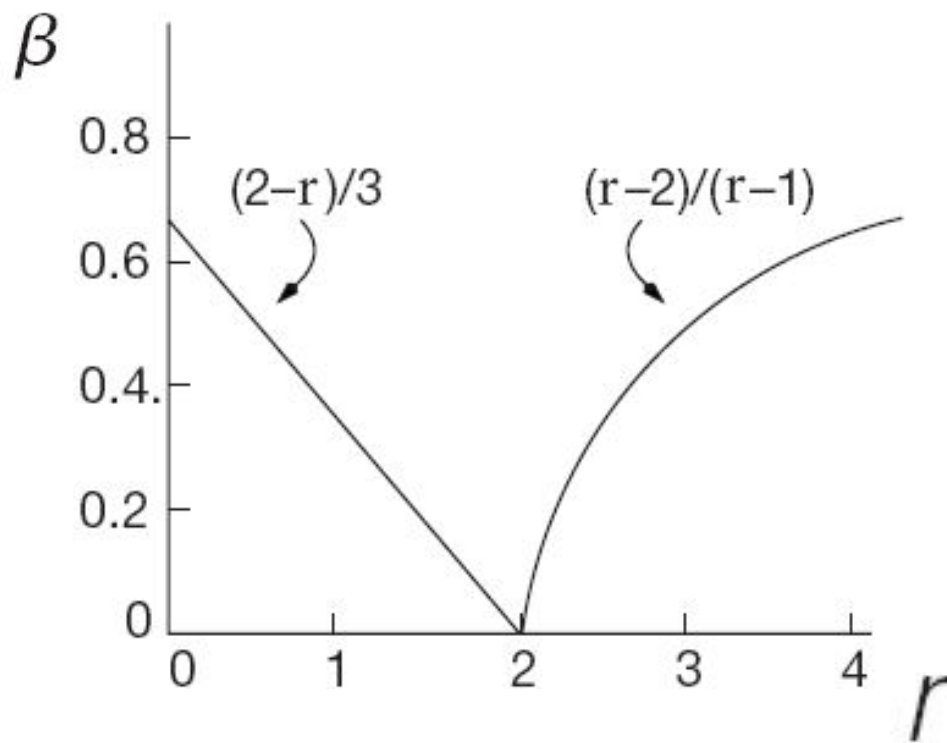
Существует константа a_r , зависящая от p, q, r , независимая от n , такая что ожидаемое время доставки любого децентрализованного алгоритма не меньше $a_r n^{(2-r)/3}$

- $r > 2$

Существует константа a_r , зависящая от p, q, r , независимая от n , такая что ожидаемое время доставки любого децентрализованного алгоритма не меньше $a_r n^{(r-2)/(r-1)}$

Зависимость ожидаемого времени от коэффициента кластеризации

- Ожидаемое время $T \geq cn^\beta$



k - мерная сеть

- обобщение результатов
- алгоритм может строить пути с длиной, полиномиально зависящей от **$\log n$** , если только **$r = k$**

Скорость передачи

- «скорость передачи» класса сетей
- минимизация диаметра не то же самое что минимизация ожидаемого времени поиска
- сеть должна содержать структурные ключи которые могут быть использованы для направления к цели

Сети, поддерживающие эффективный поиск

- Иерархическая сетевая модель
- Модель с полилогарифмическим внешним уровнем
- Модель с постоянным внешним уровнем
- Групповые структуры

Иерархическая сетевая модель

- b - арное дерево T , $b = \mathit{const}$
- V – набор листьев из T , n – размер V
- для v и w , $h(v, w)$ - высота их последнего общего предка в T
- монотонная невозрастающая функция $f(\cdot)$ - вероятность возникновения связи
- для $v \in V$, $f(h(v, w)) / \sum_{x \neq v} f(h(v, x))$
- число связей K - внешний уровень модели

Модель с полилогарифмическим внешним уровнем

- $k = c \log^2 n$, где $c = \text{const}$
- $f(h)$ растёт асимптотически как $b^{-\alpha h}$

$$\lim_{h \rightarrow \infty} \frac{f(h)}{b^{-\alpha' h}} = 0, \quad \alpha' < \alpha$$

$$\lim_{h \rightarrow \infty} \frac{b^{-\alpha'' h}}{f(h)} = 0, \quad \alpha'' > \alpha$$

Естественные интерпретации модели

- WWW иерархия тем (yahoo.com)
Science/Computer_Science/Algorithms
более вероятно будет связана с
Science/Computer_Science/Machine_Learning
, чем с ***Arts/Music/Opera***

Полученные результаты

Теорема

- \exists иерархическая модель степени $\alpha = 1$ с полилогарифмическим внешним уровнем, у которой время поиска децентрализованным алгоритмом оценивается $O(\log n)$.

Полученные результаты

Теорема(продолжение)

- $\forall \alpha \neq 1$, не существует иерархической модели степени α с полилогарифмическим внешним уровнем, у которой децентрализованный алгоритм может достичь полилогарифмическое время поиска.

Групповые структуры

- набор узлов V
- собрание подмножеств V
- константы $\lambda < 1$ и $\beta > 1$:
 - R - группа размером $q \geq 2$, $v \in R$, $\exists R' \subseteq R$, $v \in R'$, R' строго меньшая R , $|R'| \leq \lambda q$
 - R_1, R_2, R_3, \dots – группы, имеющие размер не больше q и содержащие общий узел v , то их объединение имеет размер не больше βq

Индукцированная групповая модель

Степени α

- $(V, \{R_i\})$
- $q(v, w)$ - размер наименьшей подгруппы
- $f(\cdot)$ – монотонная, невозрастающая
- $f(\cdot)$ растет асимптотически как $q^{-\alpha}$

Полученные результаты

Теорема:

- Для каждой групповой структуры существует индуцированная групповая модель степени $\alpha = 1$ с полилогарифмическим внешним уровнем, у которой время поиска децентрализованным алгоритмом оценивается $O(\log n)$.
- Для каждого $\alpha < 1$, не существует индуцированной групповой модели степени α с полилогарифмическим внешним уровнем, у которой децентрализованный алгоритм может достичь полилогарифмическое время поиска.

Выводы

- соотношение между локальной структурой и дальними контактами
- вблизи критического порога – появляются «ключи» сети.
- ниже критического значения эти ключи исчезают
- в пределе короткие цепи существуют, но индивидуумы, дезориентированные в социальных контактах, не могут их найти.

Открытые вопросы

- Вопрос Фрагно
- Какие из развивающихся процессов могут сделать поиск по сетям более эффективным?
- Осознанность узлов-посредников
- Реконструкция сетей

Ссылки

- J. Kleinberg. Navigation in a Small World. Nature 406 (2000)
- J. Kleinberg. The small-world phenomenon: An algorithmic perspective. Proc 32nd Symposium on Theory of Computing, 2000
- J. Kleinberg. Small-world Phenomena and the Dynamics of Information. Advances in Neural Information Processing Systems (NIPS) 14, 2001.

Ссылки(2)

- J. Kleinberg, P.Raghavan. Query Incentive Networks. Proc 46 th IEEE Symposium of Foundations of Computer Science, 2005.
- S. Milgram, The small world problem. Psychology Today 1 (1967)/
- J. Kleinberg. The small world phenomenon and Decentralized Search, SIAM News, Volume 37, number 3, april 2004
- Домашняя страница Джона Клайнберга.

Вопросы?

- Всем спасибо