

**Скажи мне кто твой друг и я  
скажу кто ты: выведение  
признаков профиля пользователей  
в социальной сети**

**You Are Who You Know:  
Inferring User Profiles in Online Social Networks**

*A. Mislove,  
MPI-SWS*

*B. Viswanath,  
Rice University*

*P. Druschel,  
Northeastern  
University*

*WSDM'10, February, 2010*

*Presented by  
E. Podshivalova*

# Введение

Пустые поля признаков у пользователя



Проблема



Возможно ли вывести отсутствующие признаки?

# Материалы и данные

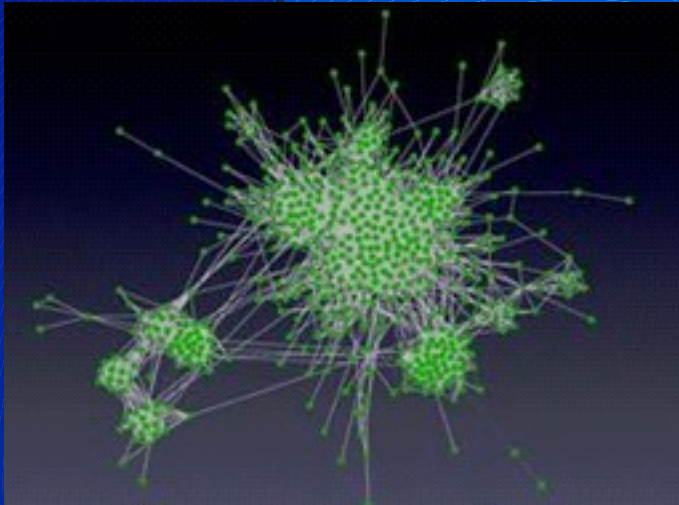
База данных студентов и аспирантов университета RICE:

*Колледж, студенческое общежитие, специализация, возраст*

+ данные открытых профилей студентов и аспирантов в Facebook (1 700 профилей): имена и отношения дружбы

Данные открытых профилей жителей Нового Орлеана в Facebook (63 000 профилей): все имеющиеся признаки

# Что такое сообщество?



Сообщество — это совокупность пользователей, которые имеют тесную связь на социальном графе.  
Связь — это отношения дружбы.

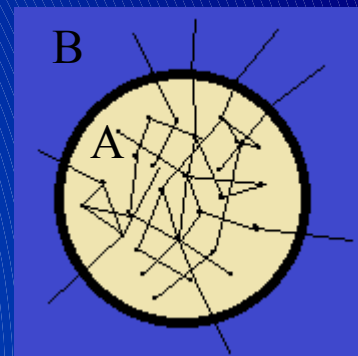
**Идея предсказания отсутствующего признака:** использовать сообщество, образовавшееся вокруг данного пользователя.

**Поиск такого сообщества:** автоматический на основании совпадения одного рассматриваемого признака и отношений дружбы.

# Метрика нормализованной проводимости (normalized conductance)

Граф  $G = (V, E)$

$$K = \frac{e_{AA}}{e_{AA} + e_{AB}} \in [0; 1]$$



$$C = \frac{e_{AA}}{e_{AA} + e_{AB}} - \frac{e_{AA} + e_{AB}}{e_{AA} + 2 \cdot e_{AB} + e_{BB}} \in [-1; 1]$$

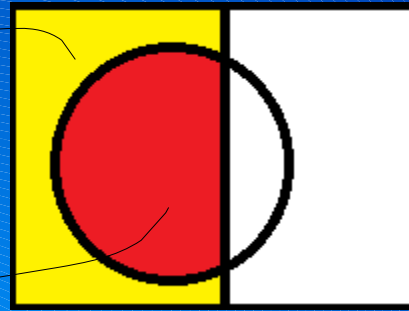
проводимость в  
случайном графе

# Алгоритм

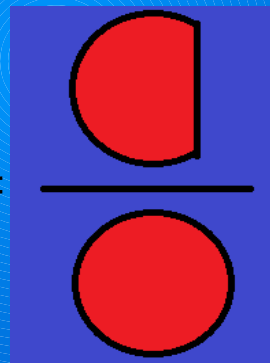
1. Дана начальная группа пользователей, имеющих один общий признак;
2. Добавление к сообществу нового пользователя, расчет метрики;
3. Пошаговая максимизация меры нормализованной проводимости;
4. Остановка алгоритма при достижении максимума метрики.

# Оценка работы алгоритма

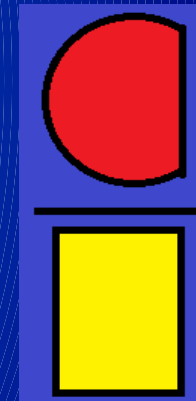
Множество  
с признаком  
Обнаружил  
алгоритм



Точность=  
(Precision)



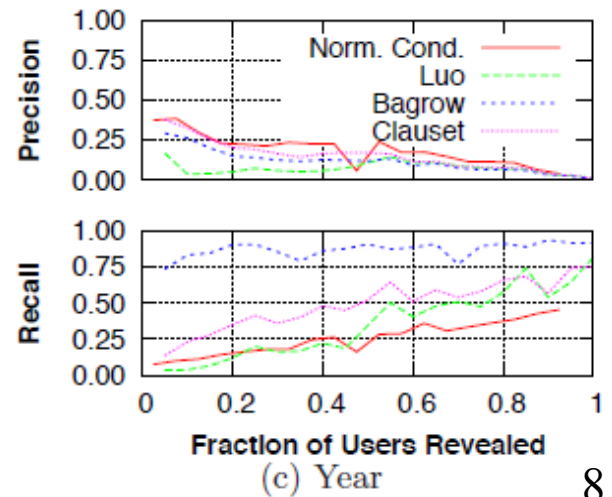
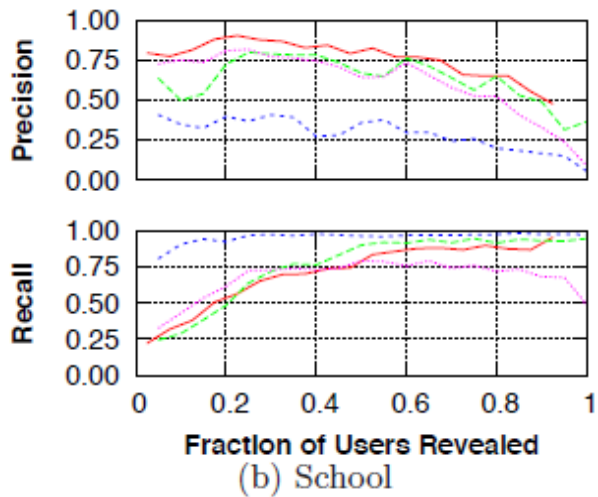
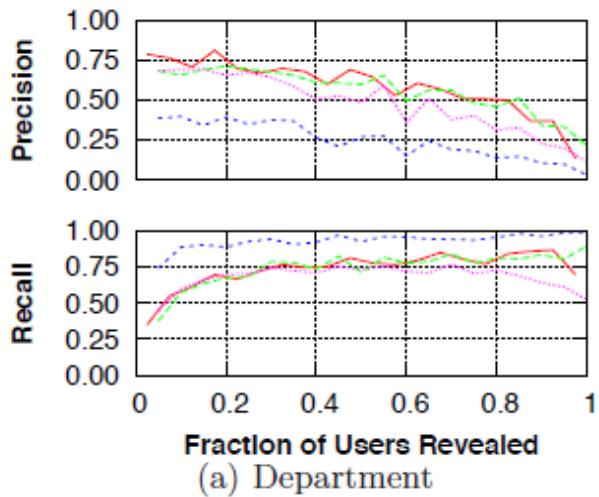
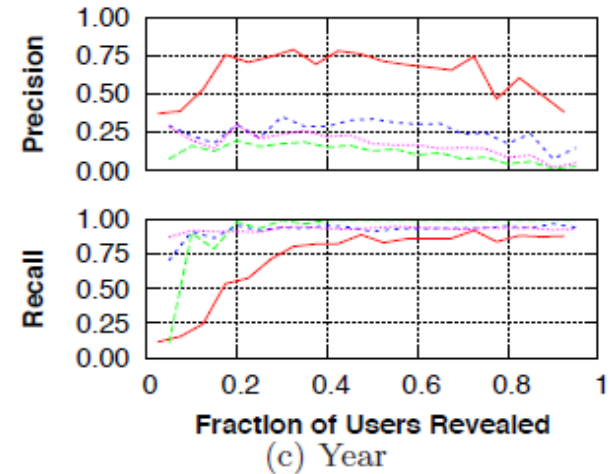
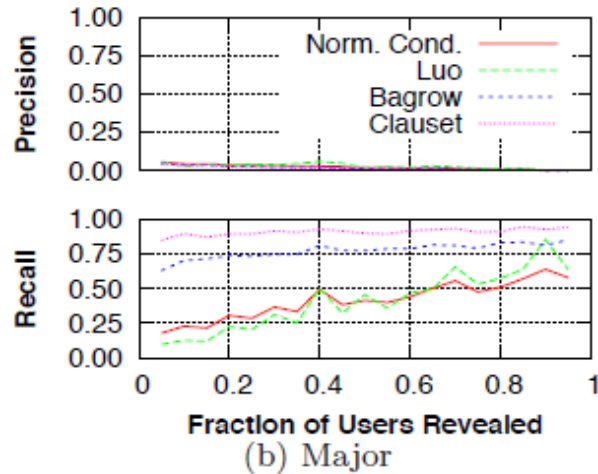
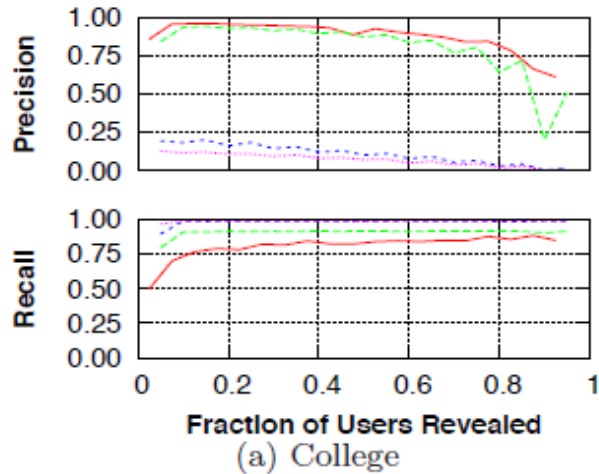
Полнота=  
(Recall)



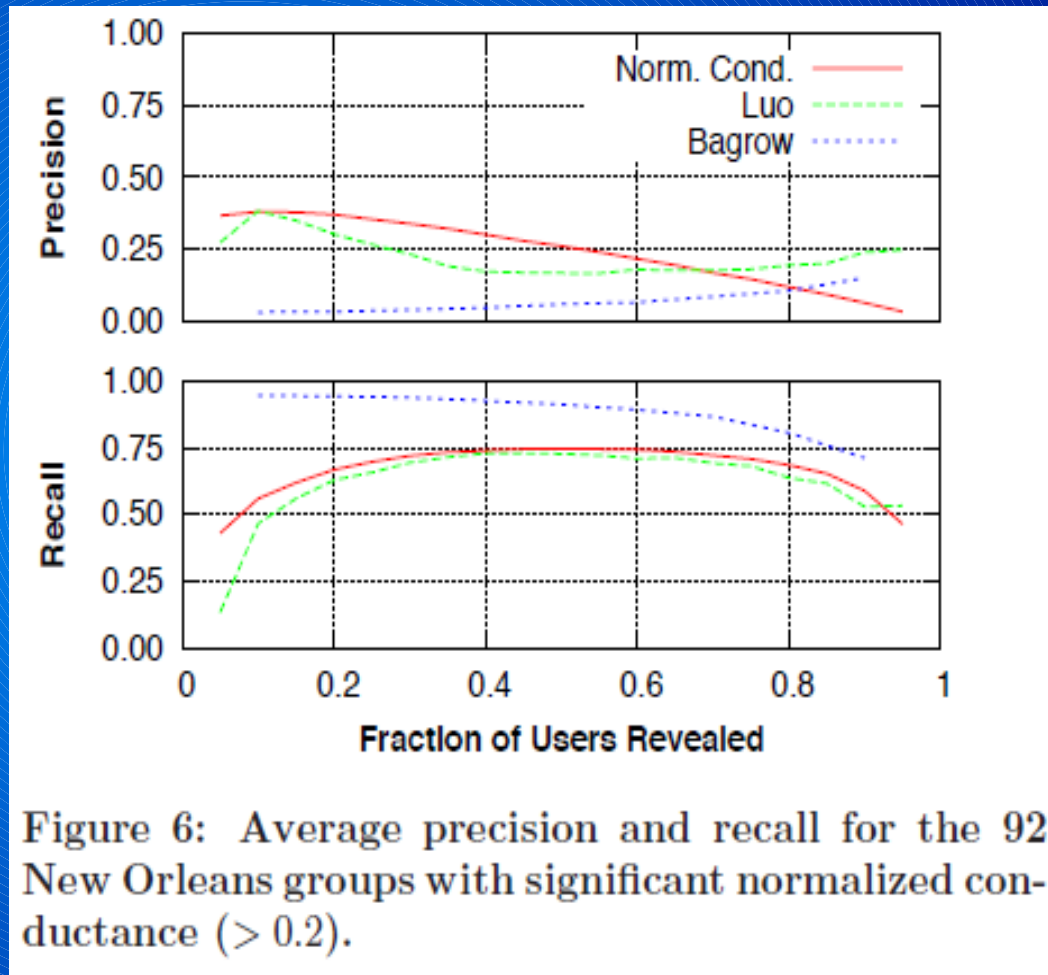
$P, R \rightarrow 1$ , если метод эффективный

$P$  и  $R$  имеют обратную связь

# Результаты работы алгоритма на данных студентов и аспирантов



# Результаты работы алгоритма на данных жителей Нового Орлеана



# Результаты

Продемонстрирован факт выводимости признаков для пользователя из признаков окружающего сообщества;

+ повышается эффективность использования социальных сетей;

+ пользователи не должны заполнять весь профиль;

- нарушение неприкосновенности частной жизни, раскрытие персональных данных.

**Спасибо за внимание!**