Случайные блуждания

Владимир Подольский

Факультет компьютерных наук, Высшая Школа Экономики

Случайные блуждания

- Задачи на графах
- Случайные блуждания
- Случайные блуждания в приложениях
- Моменты достижения из вершины
- Моменты достижения в вершину
- Блуждания длины 3

Добавление ребра

Задача

Пусть G — реальный граф из приложений. Какие ребра следует добавить к G?

- Задача звучит странно

Добавление ребра

Задача

Пусть G — реальный граф из приложений. Какие ребра следует добавить к G?

- Задача звучит странно
- Не ясно, что это вообще значит и как отвечать на такой вопрос

Добавление ребра

Задача

Пусть G — реальный граф из приложений. Какие ребра следует добавить к G?

- Задача звучит странно
- Не ясно, что это вообще значит и как отвечать на такой вопрос
- Но такие задачи реально возникают на практике

Добавление ребра

Задача

Пусть G — реальный граф из приложений. Какие ребра следует добавить к G?

- Задача звучит странно
- Не ясно, что это вообще значит и как отвечать на такой вопрос
- Но такие задачи реально возникают на практике
- Конкретная постановка зависит от контекста

Пример: соцсети

- Дан граф соцсети

Пример: соцсети

- Дан граф соцсети
- Мы хотим предложить пользователю добавить друзей

Пример: соцсети

- Дан граф соцсети
- Мы хотим предложить пользователю добавить друзей
- Хотим предложить тех, кого он скорее всего знает

Пример: соцсети

- Дан граф соцсети
- Мы хотим предложить пользователю добавить друзей
- Хотим предложить тех, кого он скорее всего знает
- По сути хотим угадать, каких ребер не хватает в графе соцсети

Пример: рекомендация покупок

- В нашем онлайн магазине есть товары

Пример: рекомендация покупок

- В нашем онлайн магазине есть товары
- Мы знаем, какие товары часто покупают вместе

Пример: рекомендация покупок

- В нашем онлайн магазине есть товары
- Мы знаем, какие товары часто покупают вместе
- Хотим понять, что еще можно предложить пользователям, которые купили какой-то товар

Пример: рекомендация покупок

- В нашем онлайн магазине есть товары
- Мы знаем, какие товары часто покупают вместе
- Хотим понять, что еще можно предложить пользователям, которые купили какой-то товар
- Граф: соединяем ребрами те товары, которые часто покупают вместе

Пример: рекомендация покупок

- В нашем онлайн магазине есть товары
- Мы знаем, какие товары часто покупают вместе
- Хотим понять, что еще можно предложить пользователям, которые купили какой-то товар
- Граф: соединяем ребрами те товары, которые часто покупают вместе
- По сути хотим угадать, каких ребер не хватает в графе

Пример: рекомендация видео

- У нас есть видео сервис, на котором пользователи просматривают видео

Пример: рекомендация видео

- У нас есть видео сервис, на котором пользователи просматривают видео
- Мы знаем всю информацию о том, кто из пользователей что посмотрел

Пример: рекомендация видео

- У нас есть видео сервис, на котором пользователи просматривают видео
- Мы знаем всю информацию о том, кто из пользователей что посмотрел
- Хотим рекомендовать пользователю новые видео

Пример: рекомендация видео

- У нас есть видео сервис, на котором пользователи просматривают видео
- Мы знаем всю информацию о том, кто из пользователей что посмотрел
- Хотим рекомендовать пользователю новые видео
- Граф: соединяем ребрами пользователя с теми видео, которые он посмотрел

Пример: рекомендация видео

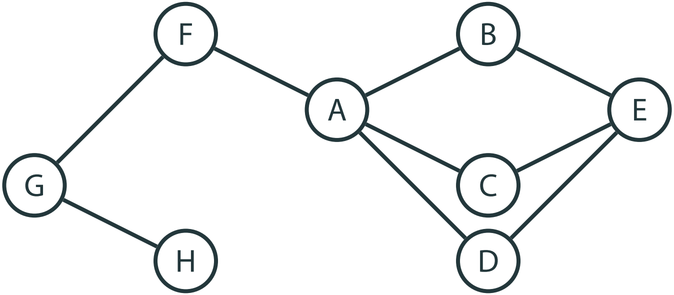
- У нас есть видео сервис, на котором пользователи просматривают видео
- Мы знаем всю информацию о том, кто из пользователей что посмотрел
- Хотим рекомендовать пользователю новые видео
- Граф: соединяем ребрами пользователя с теми видео, которые он посмотрел
- По сути хотим угадать, каких ребер не хватает в графе

Первая попытка решения

- Как же решать эту задачу?

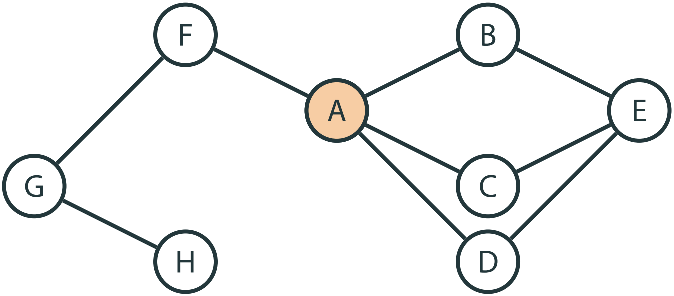
Первая попытка решения

- Как же решать эту задачу?
- Рассмотрим пример соцсети



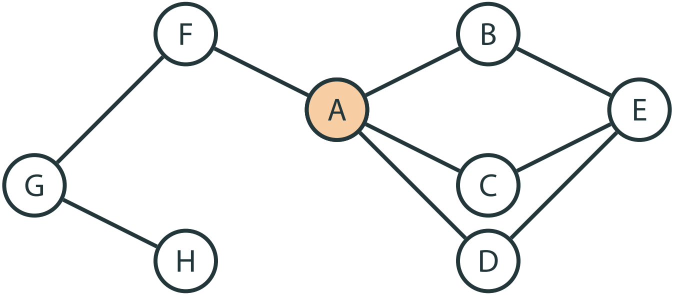
Первая попытка решения

- Как же решать эту задачу?
- Рассмотрим пример соцсети
- Кого предложить в друзья пользователю A?



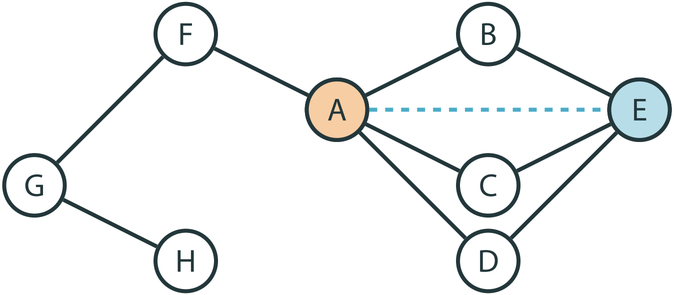
Первая попытка решения

- Идея: если у двух людей много общих друзей, они скорее всего знакомы



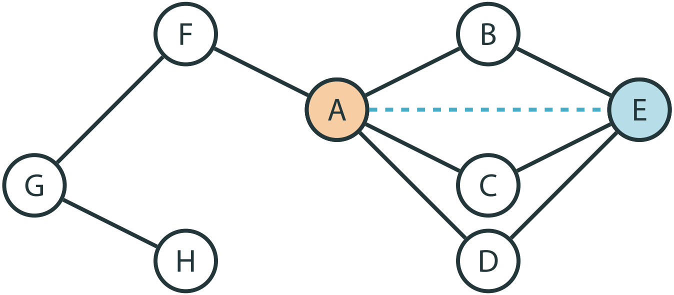
Первая попытка решения

- Идея: если у двух людей много общих друзей, они скорее всего знакомы
- Можно попробовать предлагать тех пользователей, которые с A еще не друзья, но с которыми у A больше всего общих друзей



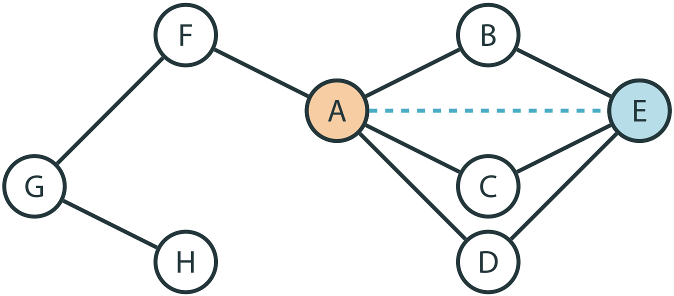
Первая попытка решения

- Оказывается, что уже этот подход хорошо работает на практике!



Первая попытка решения

- Оказывается, что уже этот подход хорошо работает на практике!
- Но мы обсудим и другие подходы



Зачем другие подходы?

- Предложения по числу общих друзей неплохо работают в соцсетях

Зачем другие подходы?

- Предложения по числу общих друзей неплохо работают в соцсетях
- Но могут плохо работать в других ситуациях

Зачем другие подходы?

- Предложения по числу общих друзей неплохо работают в соцсетях
- Но могут плохо работать в других ситуациях
- В целом, плохо работают в разреженных графах

Зачем другие подходы?

- Предложения по числу общих друзей неплохо работают в соцсетях
- Но могут плохо работать в других ситуациях
- В целом, плохо работают в разреженных графах
- Даже в соцсетях, будет странно предлагать пользователю каждый день одних и тех же друзей

Зачем другие подходы?

- Предложения по числу общих друзей неплохо работают в соцсетях
- Но могут плохо работать в других ситуациях
- В целом, плохо работают в разреженных графах
- Даже в соцсетях, будет странно предлагать пользователю каждый день одних и тех же друзей
- Нужны разные методы

Зачем другие подходы?

- Предложения по числу общих друзей неплохо работают в соцсетях
- Но могут плохо работать в других ситуациях
- В целом, плохо работают в разреженных графах
- Даже в соцсетях, будет странно предлагать пользователю каждый день одних и тех же друзей
- Нужны разные методы
- Самые эффективные и продвинутые методы часто получаются хитрой комбинацией простых методов

Зачем другие подходы?

- Предложения по числу общих друзей неплохо работают в соцсетях
- Но могут плохо работать в других ситуациях
- В целом, плохо работают в разреженных графах
- Даже в соцсетях, будет странно предлагать пользователю каждый день одних и тех же друзей
- Нужны разные методы
- Самые эффективные и продвинутые методы часто получаются хитрой комбинацией простых методов
- Полезно знать несколько простых методов

Тестирование

- У нас уже есть простой метод: число общих соседей

Тестирование

- У нас уже есть простой метод: число общих соседей
- И будут еще методы, основанные на случайных блужданиях

Тестирование

- У нас уже есть простой метод: число общих соседей
- И будут еще методы, основанные на случайных блужданиях
- Возможно они неплохо работают на практике

Тестирование

- У нас уже есть простой метод: число общих соседей
- И будут еще методы, основанные на случайных блужданиях
- Возможно они неплохо работают на практике
- Но как это проверять?

Тестирование

- Стандартный подход здесь такой

Тестирование

- Стандартный подход здесь такой
- Мы можем взять реальный граф

Тестирование

- Стандартный подход здесь такой
- Мы можем взять реальный граф
- Выберем в нем какую-то вершину

Тестирование

- Стандартный подход здесь такой
- Мы можем взять реальный граф
- Выберем в нем какую-то вершину
- Удалим несколько случайных ребер из нее

Тестирование

- Стандартный подход здесь такой
- Мы можем взять реальный граф
- Выберем в нем какую-то вершину
- Удалим несколько случайных ребер из нее
- А затем запустим наши методы на получившемся графе в этой вершине

Тестирование

- Стандартный подход здесь такой
- Мы можем взять реальный граф
- Выберем в нем какую-то вершину
- Удалим несколько случайных ребер из нее
- А затем запустим наши методы на получившемся графе в этой вершине
- Мы знаем какие в реальности надо было бы провести ребра: те, которые мы удалили!

Тестирование

- Стандартный подход здесь такой
- Мы можем взять реальный граф
- Выберем в нем какую-то вершину
- Удалим несколько случайных ребер из нее
- А затем запустим наши методы на получившемся графе в этой вершине
- Мы знаем какие в реальности надо было бы провести ребра: те, которые мы удалили!
- И мы можем проверить, насколько близкий результат покажут наши методы

Случайные блуждания

Задачи на графах

Случайные блуждания

Случайные блуждания в приложениях

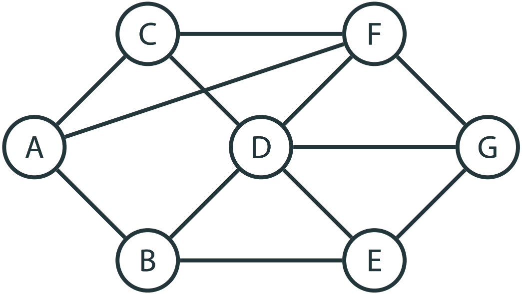
Моменты достижения из вершины

Моменты достижения в вершину

Блуждания длины 3

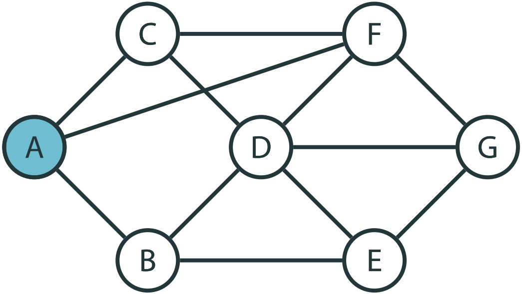
Случайные блуждания

- Пусть у нас есть граф



Случайные блуждания

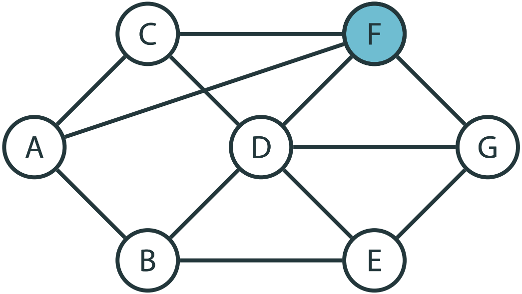
- Пусть у нас есть граф
- Рассмотрим какую-то вершину



Путь: A

Случайные блуждания

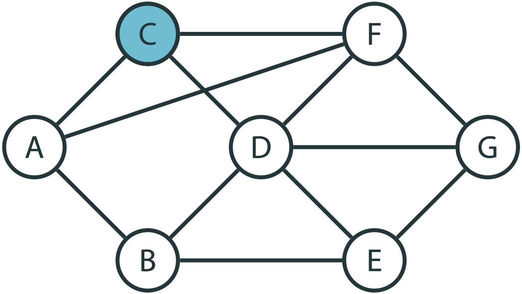
- Пусть у нас есть граф
- Рассмотрим какую-то вершину
- Перейдем случайно и равновероятно в одного из ее соседей



Путь: A, F

Случайные блуждания

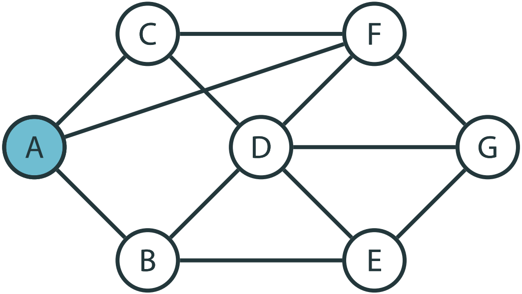
- Повторим такой переход несколько раз, каждый раз выбирая следующего соседа случайно и равновероятно



Путь: A, F, C

Случайные блуждания

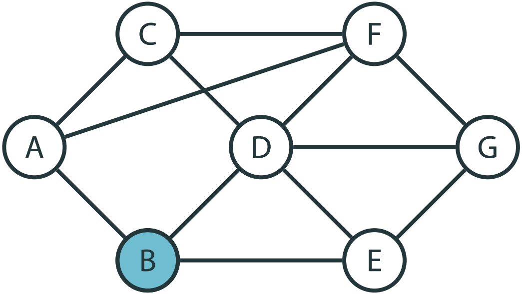
- Повторим такой переход несколько раз, каждый раз выбирая следующего соседа случайно и равновероятно



Путь: A, F, C, A

Случайные блуждания

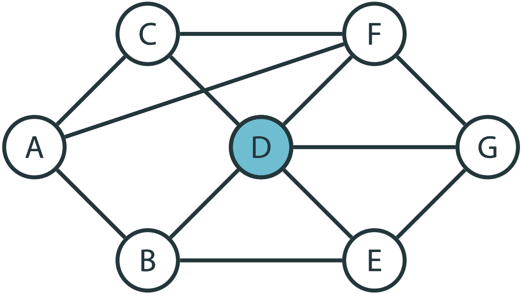
- Повторим такой переход несколько раз, каждый раз выбирая следующего соседа случайно и равновероятно



Путь: A, F, C, A, B

Случайные блуждания

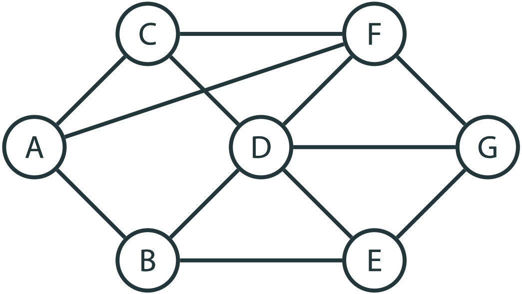
- Повторим такой переход несколько раз, каждый раз выбирая следующего соседа случайно и равновероятно



Путь: A, F, C, A, B, D

Случайные блуждания

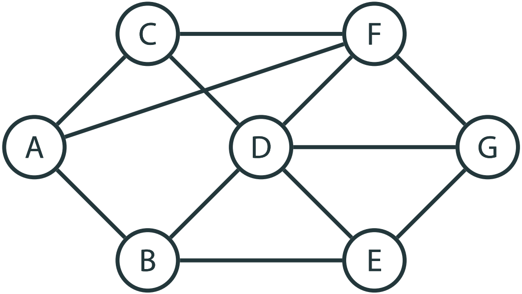
- Повторим такой переход несколько раз, каждый раз выбирая следующего соседа случайно и равновероятно
- Такой процесс называется случайным блужданием



Путь: A, F, C, A, B, D

Случайные блуждания

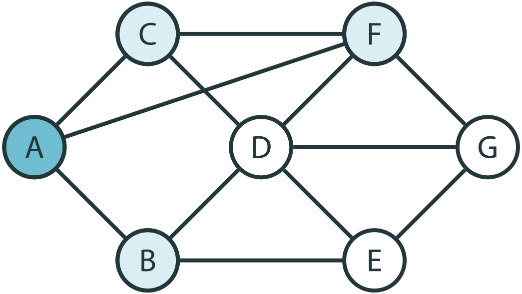
- Какова вероятность получить такой путь?



Путь: A, F, C, A, B, D

Случайные блуждания

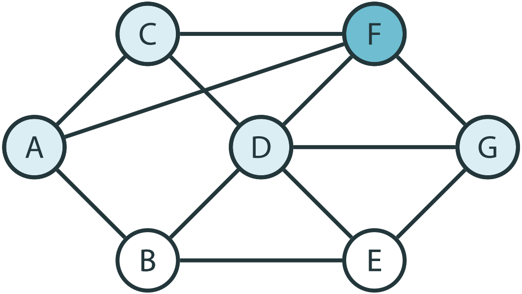
- Какова вероятность получить такой путь?
- Вероятность перейти в вершину F на первом ходу равна $\frac{1}{3}$



Путь: A, F, C, A, B, D

Случайные блуждания

- Вероятность перейти в вершину C на втором ходу равна $\frac{1}{4}$, и так далее

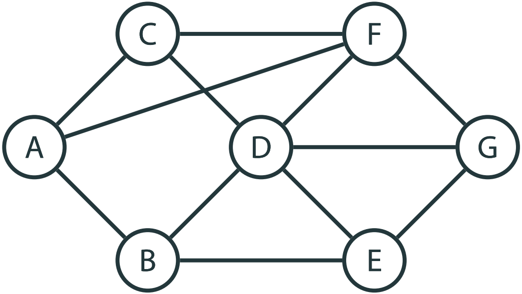


Путь: A, F, C, A, B, D

Случайные блуждания

- Вероятность перейти в вершину C на втором ходу равна $\frac{1}{4}$, и так далее

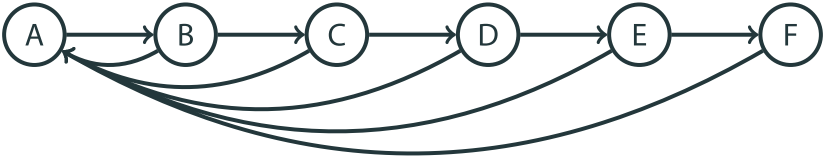
- Нужно перемножить эти вероятности:
$\frac{1}{3} \cdot \frac{1}{4} \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{324}$



Путь: A, F, C, A, B, D

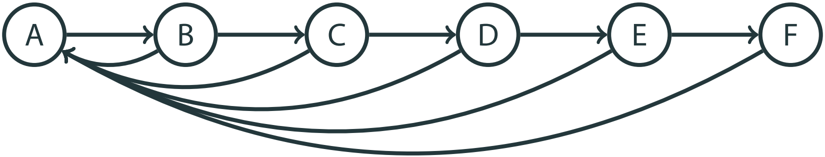
Ориентированные графы

- Блуждания можно рассматривать и в ориентированых графах



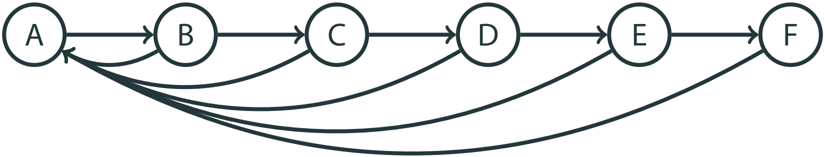
Ориентированные графы

- Блуждания можно рассматривать и в ориентированых графах
- Оказывается, что в ориентированном графе вершины могут быть труднодостижимы для случайных блужданий



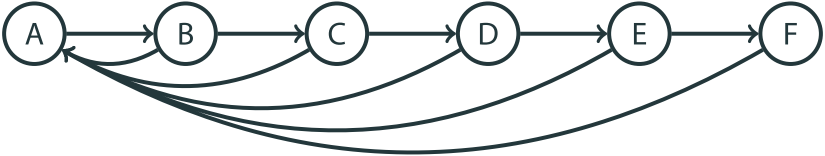
Ориентированные графы

- Блуждания можно рассматривать и в ориентированых графах
- Оказывается, что в ориентированном графе вершины могут быть труднодостижимы для случайных блужданий
- Сколько нужно ходить, чтобы попасть из A в F?



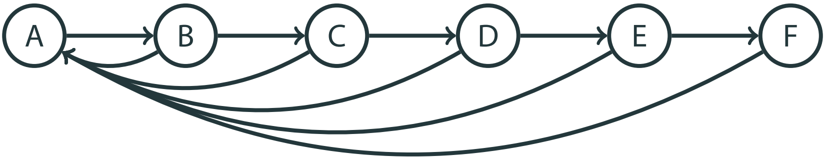
Ориентированные графы

- На каждом шаге с вероятностью $1/2$ сдвигаемся вправо



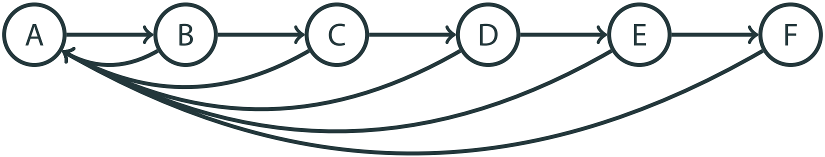
Ориентированные графы

- На каждом шаге с вероятностью $1/2$ сдвигаемся вправо
- И с вероятностью $1/2$ начинаем все сначала



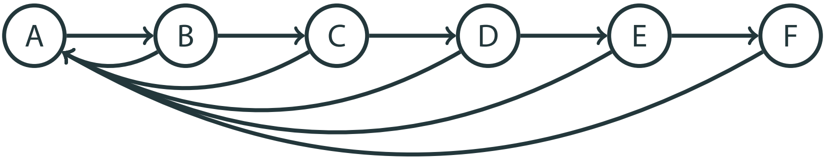
Ориентированные графы

- На каждом шаге с вероятностью $1/2$ сдвигаемся вправо
- И с вероятностью $1/2$ начинаем все сначала
- За первые 5 шагов вероятность дойти будет $\frac{1}{2^4} = \frac{1}{16}$



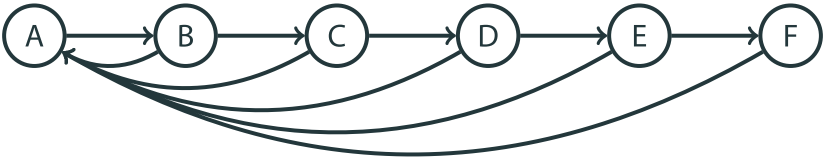
Ориентированные графы

- Если длина цепочки равна n, то вероятность дойти за n шагов будет $\frac{1}{2^{n-1}}$



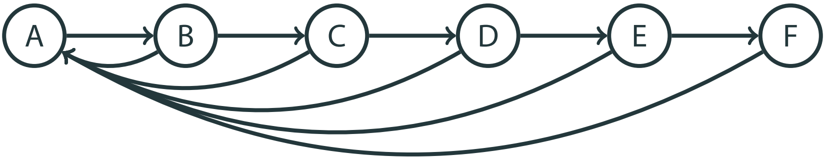
Ориентированные графы

- Если длина цепочки равна n, то вероятность дойти за n шагов будет $\frac{1}{2^{n-1}}$
- Можно доказать, что в среднем нужно порядка 2^n шагов, чтобы дойти до F



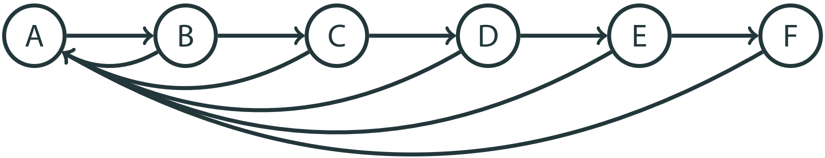
Ориентированные графы

- Если длина цепочки равна n, то вероятность дойти за n шагов будет $\frac{1}{2^{n-1}}$
- Можно доказать, что в среднем нужно порядка 2^n шагов, чтобы дойти до F
- Это очень много



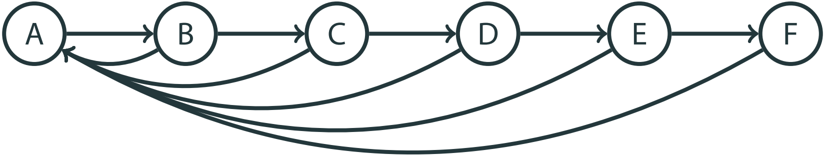
Ориентированные графы

- На практике такой плохой случай вряд ли встретится



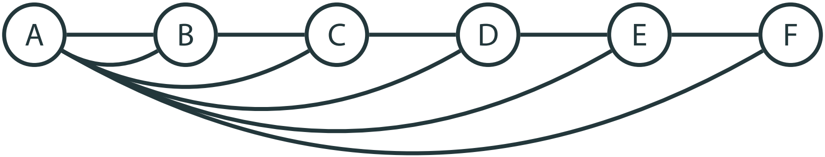
Ориентированные графы

- На практике такой плохой случай вряд ли встретится
- Но полезно помнить, что это в принципе возможно



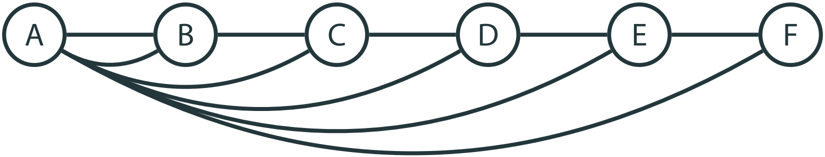
Неориентированные графы

- В неориентированных графах все гораздо лучше



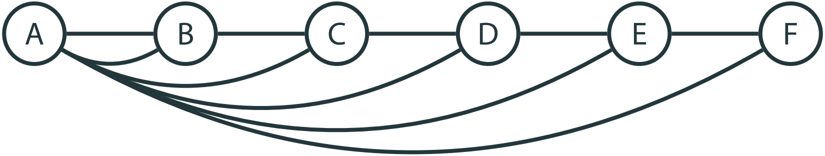
Неориентированные графы

- В неориентированных графах все гораздо лучше
- Если длина n, то из A попадаем сразу в F за один ход с вероятностью $1/n$



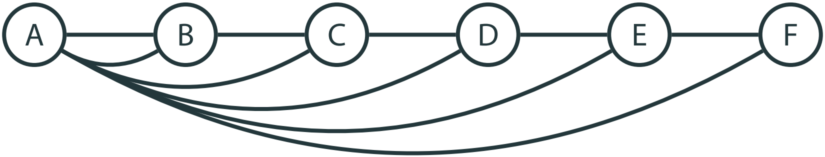
Неориентированные графы

- В неориентированных графах все гораздо лучше
- Если длина n, то из A попадаем сразу в F за один ход с вероятностью $1/n$
- Мы будем часто попадать в A



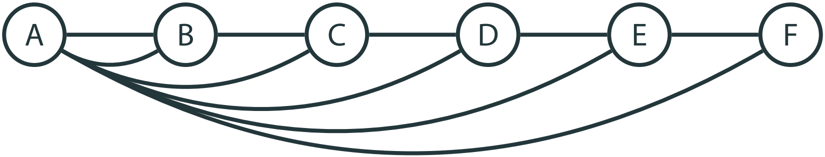
Неориентированные графы

- Можно показать, что в среднем нужно не больше $C \cdot n$ шагов, чтобы достигнуть F



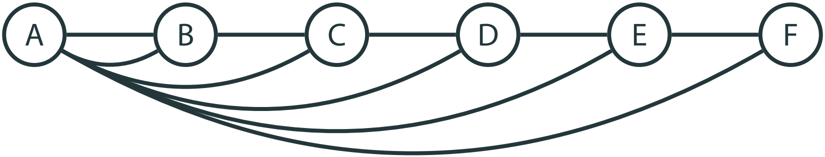
Неориентированные графы

- Можно показать, что в среднем нужно не больше $C \cdot n$ шагов, чтобы достигнуть F
- Здесь C — некоторая фиксированная константа



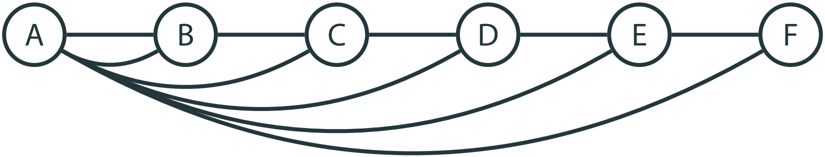
Неориентированные графы

- Важно: мы говорим только о средней длине блуждания



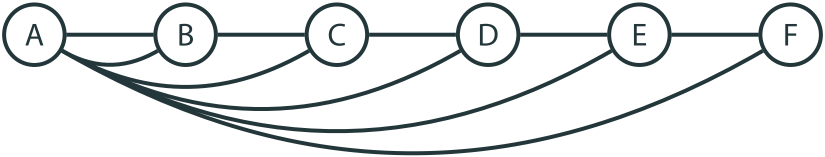
Неориентированные графы

- Важно: мы говорим только о средней длине блуждания
- Теоретически, блуждание может ходить очень долго, так и недостигнув F



Неориентированные графы

- Важно: мы говорим только о средней длине блуждания
- Теоретически, блуждание может ходить очень долго, так и недостигнув F
- Но вероятность этого очень мала, так что в среднем получается порядка n шагов



Неориентированные графы

- Такой пример несколько сложнее

Неориентированные графы

- Такой пример несколько сложнее
- В нем в среднем потребуется порядка n^2 шагов

Неориентированные графы

- Такой пример несколько сложнее
- В нем в среднем потребуется порядка n^2 шагов
- Но всегда для неориентированных связных графов на n вершинах в среднем достаточно полиномиального от n числа шагов

Неравновероятные случайные блуждания

- Мы также можем рассматривать неравновероятные случайные блуждания

Неравновероятные случайные блуждания

- Мы также можем рассматривать неравновероятные случайные блуждания
- Вероятности переходов из текущей вершины в ее соседей могут быть разными для разных соседей

Неравновероятные случайные блуждания

- Мы также можем рассматривать неравновероятные случайные блуждания
- Вероятности переходов из текущей вершины в ее соседей могут быть разными для разных соседей
- Это позволяет учитывать количественные характеристики связей между вершинами

Случайные блуждания

Задачи на графах

Случайные блуждания

Случайные блуждания в приложениях

Моменты достижения из вершины

Моменты достижения в вершину

Блуждания длины 3

Случайные блуждания

- Для чего нужны случайные блуждания?

Случайные блуждания

- Для чего нужны случайные блуждания?
- Есть много разных применений

Случайные блуждания

- Для чего нужны случайные блуждания?
- Есть много разных применений
- Мы обсудим несколько примеров

Случайные блуждания: ранжирование

- PageRank — известный алгоритм ранжирований веб-страниц

Случайные блуждания: ранжирование

- PageRank — известный алгоритм ранжирований веб-страниц
- Присваивает каждой странице число в соответствии с ее важностью

Случайные блуждания: ранжирование

- PageRank — известный алгоритм ранжирований веб-страниц
- Присваивает каждой странице число в соответствии с ее важностью
- Рассматривает граф, в котором страницы соединены ребром, если есть ссылка с одной на другую

Случайные блуждания: ранжирование

- PageRank — известный алгоритм ранжирований веб-страниц
- Присваивает каждой странице число в соответствии с ее важностью
- Рассматривает граф, в котором страницы соединены ребром, если есть ссылка с одной на другую
- По сути, вычисляет для каждой страницы вероятность посещения при последовательных случайных переходах по ссылкам

Случайные блуждания: ранжирование

- PageRank — известный алгоритм ранжирований веб-страниц
- Присваивает каждой странице число в соответствии с ее важностью
- Рассматривает граф, в котором страницы соединены ребром, если есть ссылка с одной на другую
- По сути, вычисляет для каждой страницы вероятность посещения при последовательных случайных переходах по ссылкам
- В основе — случайные блуждания

Случайные блуждания: окрестности

- Случайные блуждания помогают анализировать окрестности объектов в графах

Случайные блуждания: окрестности

- Случайные блуждания помогают анализировать окрестности объектов в графах
- Полезно, например, для кластеризации вершин

Случайные блуждания: окрестности

- Случайные блуждания помогают анализировать окрестности объектов в графах
- Полезно, например, для кластеризации вершин
- Запускаем много случайных блужданий из вершины, обычно коротких

Случайные блуждания: окрестности

- Случайные блуждания помогают анализировать окрестности объектов в графах
- Полезно, например, для кластеризации вершин
- Запускаем много случайных блужданий из вершины, обычно коротких
- Смотрим на список вершин, в которых закончились блуждания

Случайные блуждания: окрестности

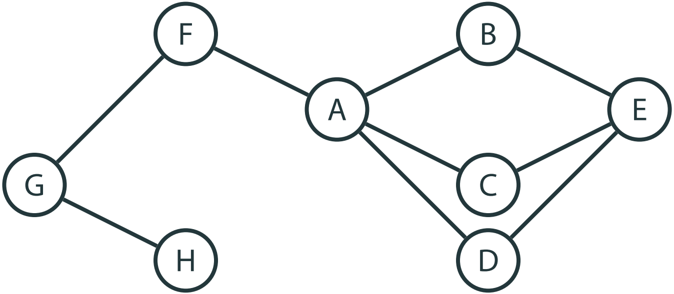
- Случайные блуждания помогают анализировать окрестности объектов в графах
- Полезно, например, для кластеризации вершин
- Запускаем много случайных блужданий из вершины, обычно коротких
- Смотрим на список вершин, в которых закончились блуждания
- Сравниваем эти списки для разных вершин

Случайные блуждания: окрестности

- Случайные блуждания помогают анализировать окрестности объектов в графах
- Полезно, например, для кластеризации вершин
- Запускаем много случайных блужданий из вершины, обычно коротких
- Смотрим на список вершин, в которых закончились блуждания
- Сравниваем эти списки для разных вершин
- Это эффективнее, чем анализировать окрестности вершин полностью

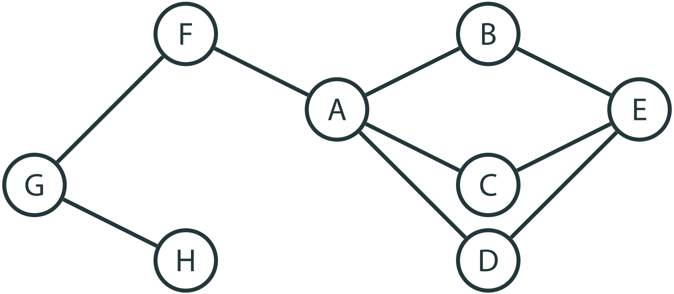
Случайные блуждания: близкие вершины

- Если есть граф, то близкие вершины это те, в которые есть короткий путь



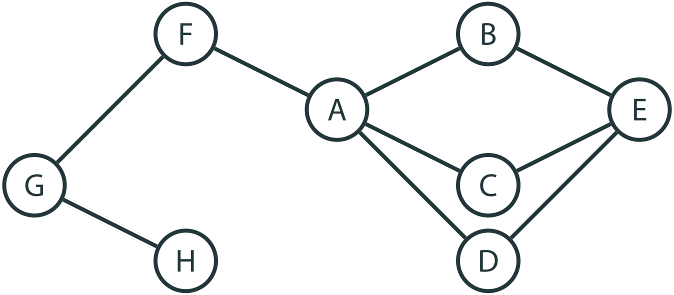
Случайные блуждания: близкие вершины

- Если есть граф, то близкие вершины это те, в которые есть короткий путь
- Но можно рассматривать другой тип близости: насколько быстро случайное блуждание приходит из одной вершины в другую



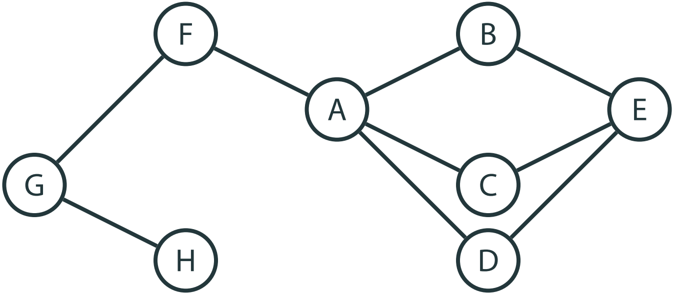
Случайные блуждания: близкие вершины

- Вершины могут быть далеки в обычном смысле



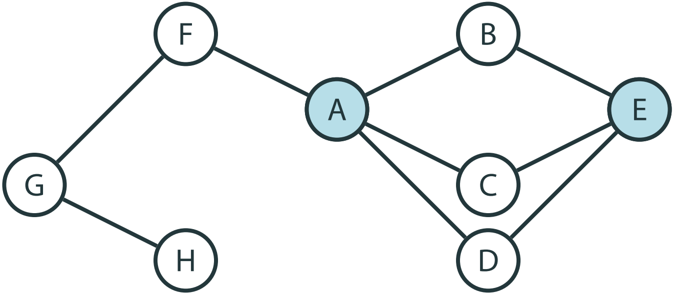
Случайные блуждания: близкие вершины

- Вершины могут быть далеки в обычном смысле
- Но если у вершин похожие окрестности, то случайное блуждание может заканчиваться быстро



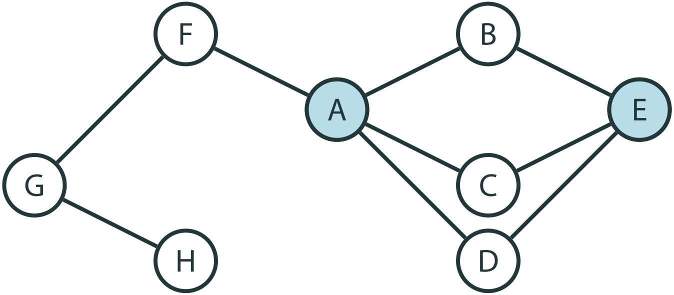
Случайные блуждания: близкие вершины

- Например, мы скорее всего попадем из A в E раньше, чем в G



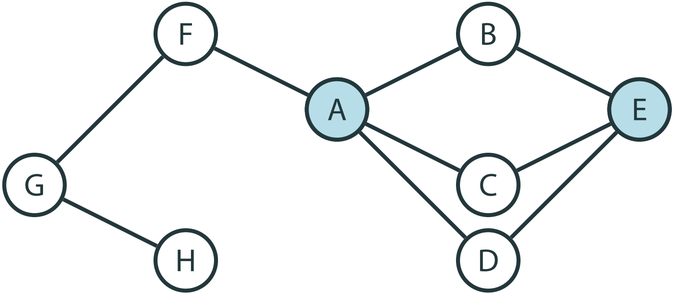
Случайные блуждания: близкие вершины

- Например, мы скорее всего попадем из A в E раньше, чем в G
- Это позволяет предсказывать, какие вершины стоит соединить в графе



Случайные блуждания: близкие вершины

- Например, мы скорее всего попадем из А в Е раньше, чем в G
- Это позволяет предсказывать, какие вершины стоит соединить в графе
- Как раз задача, которую мы обсуждали!



Момент достижения

- Пусть в графе G выбраны две вершины u и v

Момент достижения

- Пусть в графе G выбраны две вершины u и v
- Моментом достижения вершины v из вершины u называется ожидаемое число ходов случайного блуждания, стартующего из вершины u, до его попадания в вершину v

Момент достижения

- Пусть в графе G выбраны две вершины u и v
- Моментом достижения вершины v из вершины u называется ожидаемое число ходов случайного блуждания, стартующего из вершины u, до его попадания в вершину v
- Характеризует, насколько вершина v близка к вершине u

Момент достижения

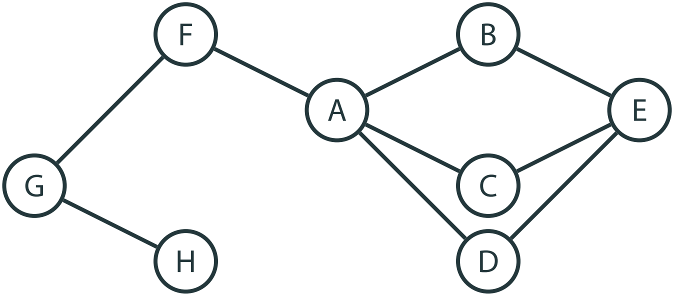
- Пусть в графе G выбраны две вершины u и v
- Моментом достижения вершины v из вершины u называется ожидаемое число ходов случайного блуждания, стартующего из вершины u, до его попадания в вершину v
- Характеризует, насколько вершина v близка к вершине u
- Минус: случайное блуждание может бесконечно долго не приходить в вершину v

Момент достижения

- Пусть в графе G выбраны две вершины u и v
- Моментом достижения вершины v из вершины u называется ожидаемое число ходов случайного блуждания, стартующего из вершины u, до его попадания в вершину v
- Характеризует, насколько вершина v близка к вершине u
- Минус: случайное блуждание может бесконечно долго не приходить в вершину v
- С этой величиной не очень удобно работать

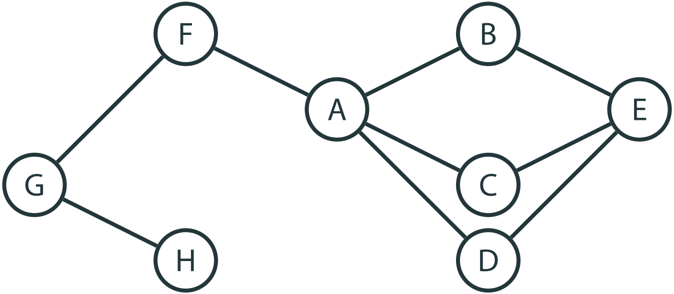
Усеченный момент достижения

- Это определение можно скорректировать



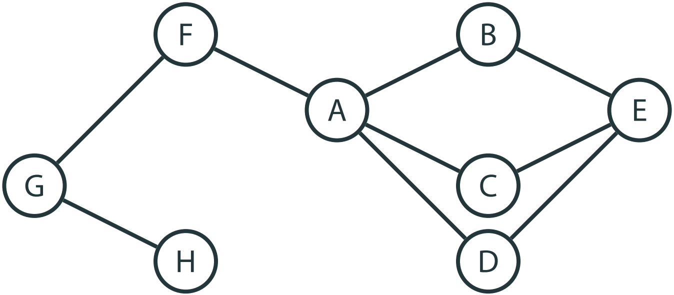
Усеченный момент достижения

- Это определение можно скорректировать
- T-усеченным моментом достижения вершины v из вершины u называется ожидаемое число ходов случайного блуждания длины T, стартующего из вершины u, до его попадания в вершину v



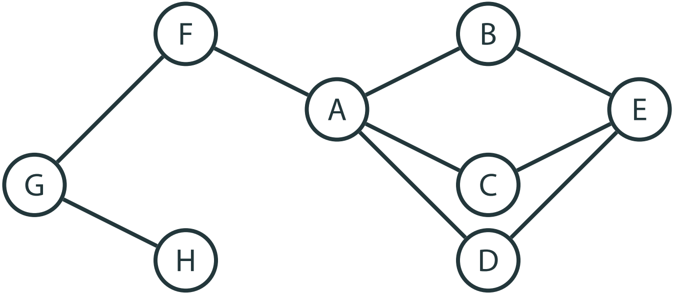
Усеченный момент достижения

- Если блуждание не попадает в вершину v за T ходов, прерываем блуждание, считаем, что сделано T ходов



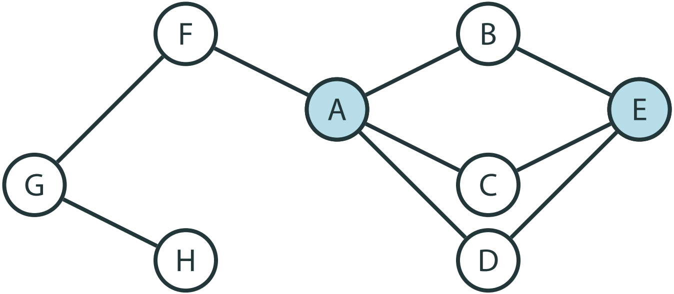
Усеченный момент достижения

- Если блуждание не попадает в вершину v за T ходов, прерываем блуждание, считаем, что сделано T ходов
- Теперь блуждание всегда конечно



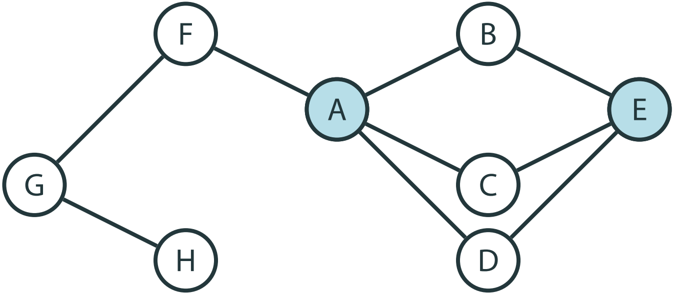
Усеченный момент достижения

- Если блуждание не попадает в вершину v за T ходов, прерываем блуждание, считаем, что сделано T ходов
- Теперь блуждание всегда конечно
- Пусть, например, мы хотим посчитать усеченный момент от A до E для $T = 3$



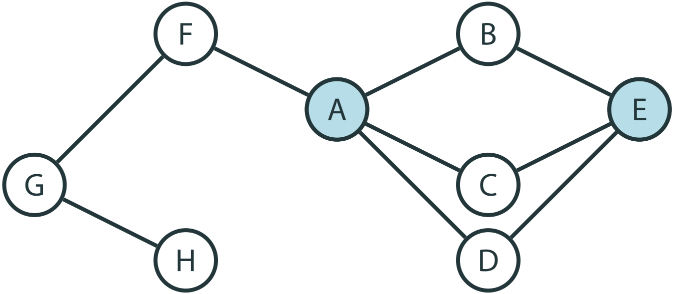
Усеченный момент достижения

- Тогда на первом шаге мы с вероятностью $\frac{3}{4}$ попадем в одну из вершин B, C или D



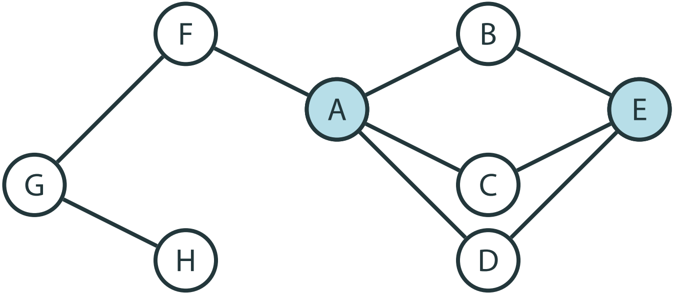
Усеченный момент достижения

- Тогда на первом шаге мы с вероятностью $\frac{3}{4}$ попадем в одну из вершин B, C или D
- И тогда на втором шаге мы с вероятностью $\frac{1}{2}$ попадем в вершину E



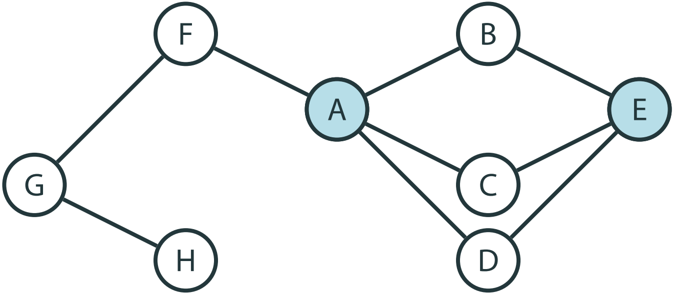
Усеченный момент достижения

- Во всех остальных случаях мы не дойдем до E быстрее чем за 3 шага



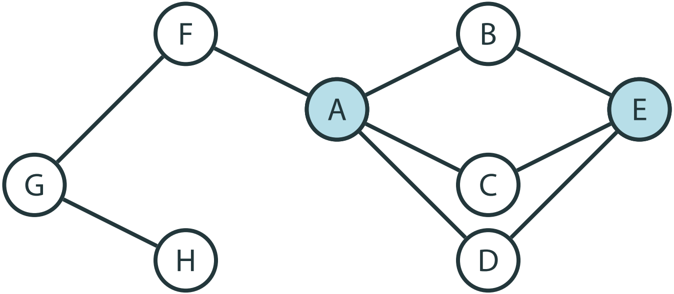
Усеченный момент достижения

- Во всех остальных случаях мы не дойдем до E быстрее чем за 3 шага
- Так что мы дойдем за 2 шага с вероятностью $\frac{3}{4} \cdot \frac{1}{2} = \frac{3}{8}$



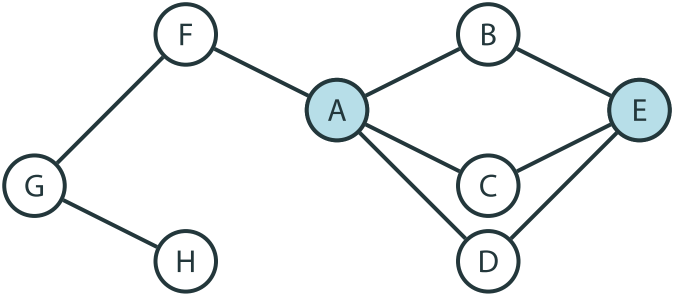
Усеченный момент достижения

- Во всех остальных случаях мы не дойдем до E быстрее чем за 3 шага
- Так что мы дойдем за 2 шага с вероятностью $\frac{3}{4} \cdot \frac{1}{2} = \frac{3}{8}$
- Тогда среднее число шагов равно $2 \cdot \frac{3}{8} + 3 \cdot \frac{5}{8} = \frac{21}{8} = 2.625$



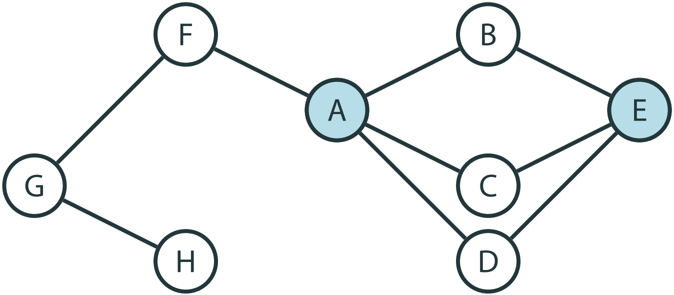
Усеченный момент достижения

- Величина несколько хуже характеризует близость вершин



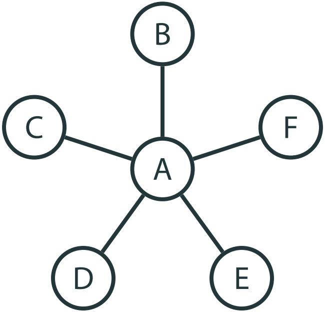
Усеченный момент достижения

- Величина несколько хуже характеризует близость вершин
- Но с ней гораздо удобнее работать



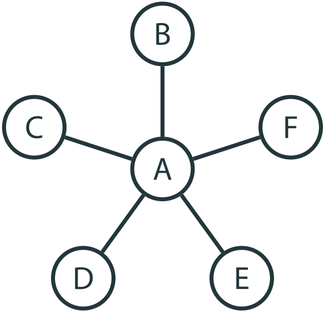
Несимметричность

- Рассматриваемая величина несимметрична: моменты достижения из v в u и из u в v могут существенно различаться



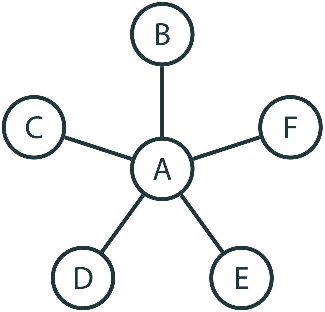
Несимметричность

- Рассматриваемая величина несимметрична: моменты достижения из v в u и из u в v могут существенно различаться
- Посчитаем момент из A в B для $T = 5$



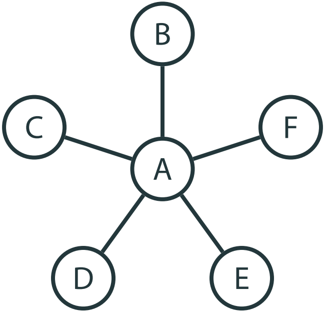
Несимметричность

- Вероятность попасть в B на 1 шаге равна $1/5$



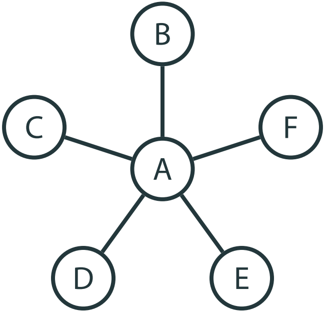
Несимметричность

- Вероятность попасть в B на 1 шаге равна $1/5$
- С вероятностью $4/5$ через два шага вернемся в A



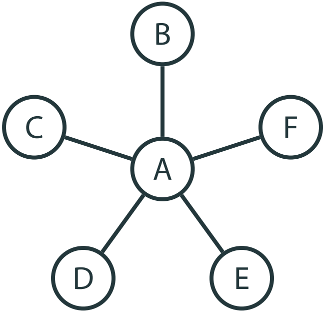
Несимметричность

- Вероятность попасть в B на 1 шаге равна $1/5$
- С вероятностью $4/5$ через два шага вернемся в A
- Вероятность попасть в B за 3 шага равна $\frac{4}{5} \cdot \frac{1}{5} = \frac{4}{25}$



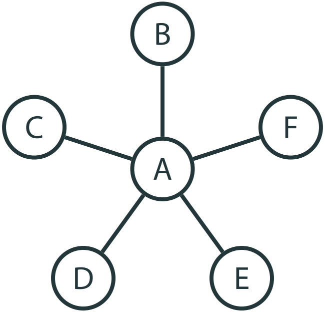
Несимметричность

- Во всех остальных случаях придется сделать 5 шагов



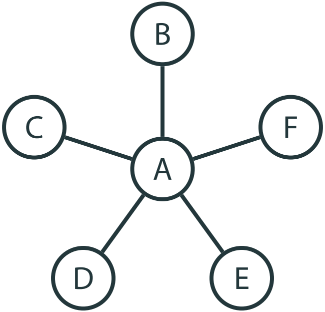
Несимметричность

- Во всех остальных случаях придется сделать 5 шагов
- Получаем среднюю длину пути
  $\frac{1}{5} + 3 \cdot \frac{4}{25} + 5 \cdot \frac{16}{25} = 3.88$



Несимметричность

- Во всех остальных случаях придется сделать 5 шагов
- Получаем среднюю длину пути
  $\frac{1}{5} + 3 \cdot \frac{4}{25} + 5 \cdot \frac{16}{25} = 3.88$
- Из B в A мы с вероятностью 1 дойдем за один шаг



Симметризация

- Но если мы хотим измерять близость вершин, может быть естественно иметь симметричную меру

Симметризация

- Но если мы хотим измерять близость вершин, может быть естественно иметь симметричную меру
- Временем перемещения между вершинами u и v называется сумма моментов достижения из u в v и из v в u

Симметризация

- Но если мы хотим измерять близость вершин, может быть естественно иметь симметричную меру
- Временем перемещения между вершинами u и v называется сумма моментов достижения из u в v и из v в u
- Аналогично T-учесенным временем перемещения между вершинами u и v называется сумма T-усеченных моментов достижения из u в v и из v в u

Симметризация

- Но если мы хотим измерять близость вершин, может быть естественно иметь симметричную меру
- Временем перемещения между вершинами u и v называется сумма моментов достижения из u в v и из v в u
- Аналогично T-учесенным временем перемещения между вершинами u и v называется сумма T-усеченных моментов достижения из u в v и из v в u
- Эта мера уже является симметричной

Что мы получили

- Итак, у нас есть три метода нахождения похожих вершин

Что мы получили

- Итак, у нас есть три метода нахождения похожих вершин
    1. Усеченные моменты из вершины

Что мы получили

- Итак, у нас есть три метода нахождения похожих вершин
    1. Усеченные моменты из вершины
    2. Усеченные моменты в вершину

Что мы получили

- Итак, у нас есть три метода нахождения похожих вершин

  1. Усеченные моменты из вершины
  2. Усеченные моменты в вершину
  3. Усеченное время перемещения

Что мы получили

- Итак, у нас есть три метода нахождения похожих вершин

    1. Усеченные моменты из вершины
    2. Усеченные моменты в вершину
    3. Усеченное время перемещения

- Для данной вершины мы можем посчитать близость ко всем остальным вершинам

Что мы получили

- Итак, у нас есть три метода нахождения похожих вершин
    1. Усеченные моменты из вершины
    2. Усеченные моменты в вершину
    3. Усеченное время перемещения
- Для данной вершины мы можем посчитать близость ко всем остальным вершинам
- Тогда можно предложить добавить ребра в самые близкие вершины

Случайные блуждания

Задачи на графах

Случайные блуждания

Случайные блуждания в приложениях

Моменты достижения из вершины

Моменты достижения в вершину

Блуждания длины 3

Подсчет моментов достижения

- Нам требуется подсчитывать усеченные моменты достижения

Подсчет моментов достижения

- Нам требуется подсчитывать усеченные моменты достижения
- Как же это делать?

Подсчет моментов достижения

- Нам требуется подсчитывать усеченные моменты достижения
- Как же это делать?
- Нас интересует задача нахождения ближайших вершин к данной

Подсчет моментов достижения

- Нам требуется подсчитывать усеченные моменты достижения
- Как же это делать?
- Нас интересует задача нахождения ближайших вершин к данной
- Так что нам нужно подсчитывать усеченные моменты из данной вершины до всех остальных

Подсчет моментов достижения

- Нам требуется подсчитывать усеченные моменты достижения
- Как же это делать?
- Нас интересует задача нахождения ближайших вершин к данной
- Так что нам нужно подсчитывать усеченные моменты из данной вершины до всех остальных
- А также усеченные моменты из всех вершин в данную

Моменты достижения из вершины

- Как подсчитывать моменты достижения из данной вершины?

Моменты достижения из вершины

- Как подсчитывать моменты достижения из данной вершины?
- По определению нужно перебрать все пути и для каждой вершины усреднить расстояние до нее по этим путям

Моменты достижения из вершины

- Как подсчитывать моменты достижения из данной вершины?
- По определению нужно перебрать все пути и для каждой вершины усреднить расстояние до нее по этим путям
- Это очень долго

Моменты достижения из вершины

- Как подсчитывать моменты достижения из данной вершины?
- По определению нужно перебрать все пути и для каждой вершины усреднить расстояние до нее по этим путям
- Это очень долго
- Вместо этого можно посчитать моменты приближенно — семплирование

Семплирование

- Запускаем S случайных блужданий из вершины

Семплирование

- Запускаем S случайных блужданий из вершины
- По каждому блужданию меряем расстояние до всех вершин

Семплирование

- Запускаем S случайных блужданий из вершины
- По каждому блужданию меряем расстояние до всех вершин
- Если T сильно меньше числа вершин, то для большинства вершин расстояние будет T

Семплирование

- Запускаем S случайных блужданий из вершины
- По каждому блужданию меряем расстояние до всех вершин
- Если T сильно меньше числа вершин, то для большинства вершин расстояние будет T
- Для каждой вершины усредняем расстояния по сделанным блужданиям

Семплирование

- Запускаем S случайных блужданий из вершины
- По каждому блужданию меряем расстояние до всех вершин
- Если T сильно меньше числа вершин, то для большинства вершин расстояние будет T
- Для каждой вершины усредняем расстояния по сделанным блужданиям
- Поскольку нас интересуют только ближайшие вершины к изначальной, не страшно, что большинство расстояний будет равно T

Моменты достижения из вершины

- Сколько времени это занимает?

Моменты достижения из вершины

- Сколько времени это занимает?
- Заводим массив по всем вершинам

A	B	C	D	E	F	G

$T = 4, S = 10$

Моменты достижения из вершины

- Сколько времени это занимает?
- Заводим массив по всем вершинам
- Проходим по случайному блужданию

A	B	C	D	E	F	G

$T = 4$, $S = 10$

Блуждание: A, C, E, C, B

Моменты достижения из вершины

- Для всех вершин в блуждании помечаем в массиве первый момент, когда они были достигнуты

A	B	C	D	E	F	G
0	4	1		2		

T = 4, S = 10

Блуждание: A, C, E, C, B

Моменты достижения из вершины

- Для всех вершин в блуждании помечаем в массиве первый момент, когда они были достигнуты
- Для остальных вершин полагаем момент равным T

A	B	C	D	E	F	G
0	4	1	4	2	4	4

T = 4, S = 10

Блуждание: A, C, E, C, B

Моменты достижения из вершины

- Для всех вершин в блуждании помечаем в массиве первый момент, когда они были достигнуты
- Для остальных вершин полагаем момент равным T
- Повторяем для каждого блуждания, усредняем

A	B	C	D	E	F	G
0	2.4	2.5	3.5	3.3	3.3	4

T = 4, S = 10

Блуждание: A, C, E, C, B

Моменты достижения из вершины

- На одно блуждание требуется порядка $T + V$ операций

A	B	C	D	E	F	G
0	2.4	2.5	3.5	3.3	3.3	4

T = 4, S = 10

Блуждание: A, C, E, C, B

Моменты достижения из вершины

- На одно блуждание требуется порядка $T + V$ операций
- Всего требуется порядка $S \cdot (T + V)$ операций

A	B	C	D	E	F	G
0	2.4	2.5	3.5	3.3	3.3	4

T = 4, S = 10

Блуждание: A, C, E, C, B

Моменты достижения из вершины

- На одно блуждание требуется порядка $T + V$ операций
- Всего требуется порядка $S \cdot (T + V)$ операций
- Обычно T сильно меньше V, так что порядка $S \cdot V$

A	B	C	D	E	F	G
0	2.4	2.5	3.5	3.3	3.3	4

T = 4, S = 10

Блуждание: A, C, E, C, B

Моменты достижения из вершины

- Можно еще эффективнее

Моменты достижения из вершины

- Можно еще эффективнее
- Заводим массив для вершин, где будем хранить сумму расстояний по всем блужданиям

	A	B	C	D	E	F	G
dist	0	0	0	0	0	0	0

$T = 4$, $S = 10$

Моменты достижения из вершины

- Можно еще эффективнее
- Заводим массив для вершин, где будем хранить сумму расстояний по всем блужданиям
- Заводим массив счетчиков для вершин

	A	B	C	D	E	F	G
dist	0	0	0	0	0	0	0
count	0	0	0	0	0	0	0

T = 4, S = 10

Моменты достижения из вершины

- Для каждого блуждания проходим по вершинам блуждания

	A	B	C	D	E	F	G
dist	0	0	0	0	0	0	0
count	0	0	0	0	0	0	0

T = 4, S = 10

Блуждание: A, C, E, C, B

Моменты достижения из вершины

- Для каждого блуждания проходим по вершинам блуждания
- Проверяем, встречалась ли данная вершина раньше в этом блуждании

	A	B	C	D	E	F	G
dist	0	0	0	0	0	0	0
count	0	0	0	0	0	0	0

T = 4, S = 10

Блуждание: A, C, E, C, B

Моменты достижения из вершины

- Если нет, то прибавляем ее номер в соответствующую ячейку массива расстояний

	A	B	C	D	E	F	G
dist	0	4	1	0	2	0	0
count	0	0	0	0	0	0	0

T = 4, S = 10

Блуждание: A, C, E, C, B

Моменты достижения из вершины

- Если нет, то прибавляем ее номер в соответствующую ячейку массива расстояний
- Увеличиваем счетчик вершины на 1

	A	B	C	D	E	F	G
dist	0	4	1	0	2	0	0
count	1	1	1	0	1	0	0

T = 4, S = 10

Блуждание: A, C, E, C, B

Моменты достижения из вершины

- Если нет, то прибавляем ее номер в соответствующую ячейку массива расстояний
- Увеличиваем счетчик вершины на 1
- Повторяем так по всем блужданиям

	A	B	C	D	E	F	G
dist	0	16	13	15	21	5	4
count	10	8	7	5	7	3	1

T = 4, S = 10

Моменты достижения из вершины

- На данный момент в массиве расстояния были учтены только расстояния для вершин, которые достигались блужданиями

	A	B	C	D	E	F	G
dist	0	16	13	15	21	5	4
count	10	8	7	5	7	3	1

$T = 4, S = 10$

Моменты достижения из вершины

- На данный момент в массиве расстояния были учтены только расстояния для вершин, которые достигались блужданиями
- Но у нас для каждой вершины есть счетчик, сколько раз это произошло

	A	B	C	D	E	F	G
dist	0	16	13	15	21	5	4
count	10	8	7	5	7	3	1

T = 4, S = 10

Моменты достижения из вершины

- Проходим по массиву расстояний и для каждой вершины прибавляем к ячейке $T \cdot (S - i)$, где i — значение счетчика для соответствующей вершины

	A	B	C	D	E	F	G
dist	0	16	13	15	21	5	4
count	10	8	7	5	7	3	1

T = 4, S = 10

Моменты достижения из вершины

- Проходим по массиву расстояний и для каждой вершины прибавляем к ячейке $T \cdot (S - i)$, где i — значение счетчика для соответствующей вершины

	A	B	C	D	E	F	G
dist	0	24	25	35	33	33	40
count	10	10	10	10	10	10	10

T = 4, S = 10

Моменты достижения из вершины

- Проходим по массиву расстояний и для каждой вершины прибавляем к ячейке $T \cdot (S - i)$, где i — значение счетчика для соответствующей вершины

- Еще раз проходим по массиву вершин и делим все результаты на S

	A	B	C	D	E	F	G
dist	0	24	25	35	33	33	40
count	10	10	10	10	10	10	10

T = 4, S = 10

Моменты достижения из вершины

- Проходим по массиву расстояний и для каждой вершины прибавляем к ячейке $T \cdot (S - i)$, где i — значение счетчика для соответствующей вершины
- Еще раз проходим по массиву вершин и делим все результаты на S

	A	B	C	D	E	F	G
dist	0	2.4	2.5	3.5	3.3	3.3	4
count	10	10	10	10	10	10	10

T = 4, S = 10

Моменты достижения из вершины

- В чем плюс: не нужно проходить по всему массиву вершин на каждом блуждании

	A	B	C	D	E	F	G
dist	0	2.4	2.5	3.5	3.3	3.3	4
count	10	10	10	10	10	10	10

T = 4, S = 10

Моменты достижения из вершины

- В чем плюс: не нужно проходить по всему массиву вершин на каждом блуждании
- Работаем со всем массивом только несколько раз: в начале и в конце

	A	B	C	D	E	F	G
dist	0	2.4	2.5	3.5	3.3	3.3	4
count	10	10	10	10	10	10	10

$T = 4, S = 10$

Моменты достижения из вершины

- В чем плюс: не нужно проходить по всему массиву вершин на каждом блуждании
- Работаем со всем массивом только несколько раз: в начале и в конце
- Потребуется порядка $V + S \cdot T$ операций

	A	B	C	D	E	F	G
dist	0	2.4	2.5	3.5	3.3	3.3	4
count	10	10	10	10	10	10	10

T = 4, S = 10

Случайные блуждания

Задачи на графах

Случайные блуждания

Случайные блуждания в приложениях

Моменты достижения из вершины

Моменты достижения в вершину

Блуждания длины 3

Моменты достижения в вершину

- Посчитали моменты достижения из вершины

Моменты достижения в вершину

- Посчитали моменты достижения из вершины
- Но как посчитать все моменты достижения в вершину

Моменты достижения в вершину

- Посчитали моменты достижения из вершины
- Но как посчитать все моменты достижения в вершину
- Прошлый подход работает плохо: случайные блуждания стартуют из разных вершин

Моменты достижения в вершину

- Посчитали моменты достижения из вершины
- Но как посчитать все моменты достижения в вершину
- Прошлый подход работает плохо: случайные блуждания стартуют из разных вершин
- Пришлось бы делать отдельные блуждания для каждой вершины

Моменты достижения в вершину

- Посчитали моменты достижения из вершины
- Но как посчитать все моменты достижения в вершину
- Прошлый подход работает плохо: случайные блуждания стартуют из разных вершин
- Пришлось бы делать отдельные блуждания для каждой вершины
- Это очень долго

Моменты достижения в вершину

- Оказывается можно посчитать точные значения усеченных моментов достижения рекурсивно!

Моменты достижения в вершину

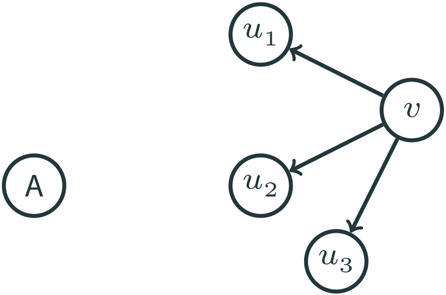
- Оказывается можно посчитать точные значения усеченных моментов достижения рекурсивно!
- Пусть $h(v, T)$ — это T-усеченный момент достижения нашей вершины из вершины v

Моменты достижения в вершину

- Во-первых, $h(v, 0) = 0$ для всех v

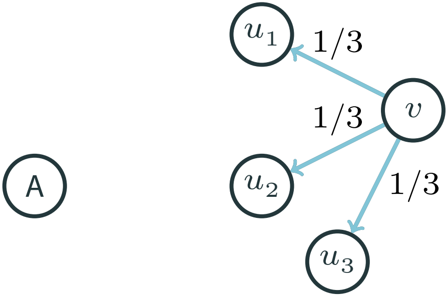
Моменты достижения в вершину

- Во-первых, $h(v, 0) = 0$ для всех v
- Для $T > 0$ посмотрим на первый шаг блуждания



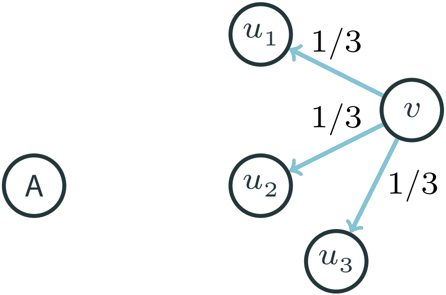
Моменты достижения в вершину

- Во-первых, $h(v, 0) = 0$ для всех v
- Для $T > 0$ посмотрим на первый шаг блуждания
- Мы переходим в одного из соседей v равновероятно



Моменты достижения в вершину

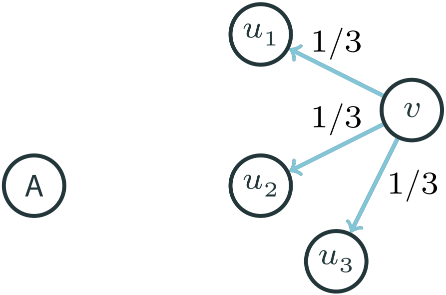
- А дальше из каждого соседа запускаем блуждание длины $T-1$!



Моменты достижения в вершину

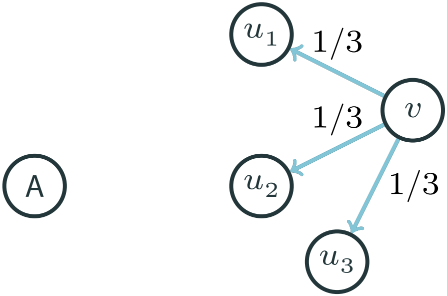
- А дальше из каждого соседа запускаем блуждание длины $T-1$!

- Ожидаемая длина случайного блуждания из v длины T равна 1 плюс усредненная ожидаемая длина блужданий длины $T-1$ по всем соседям v



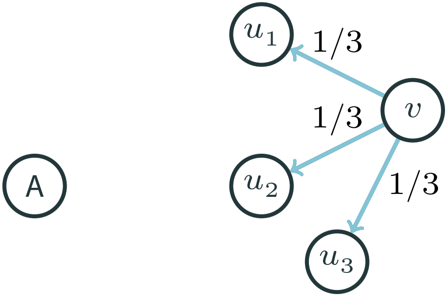
Моменты достижения в вершину

- Или $h(v,T) = 1 + \frac{1}{|N(v)|}\sum_{u\in N(v)} h(u, T-1)$



Моменты достижения в вершину

- Или $h(v,T) = 1 + \frac{1}{|N(v)|} \sum_{u \in N(v)} h(u, T-1)$
- Если мы уже посчитали $h(u, T-1)$ для всех u, то мы можем посчитать $h(v,T)$ для всех v!



Моменты достижения в вершину

- Как это реализуется?

Моменты достижения в вершину

- Как это реализуется?
- Храним двумерный массив: для каждой вершины v и каждой длины блуждания T храним $h(v, T)$

Моменты достижения в вершину

- Как это реализуется?
- Храним двумерный массив: для каждой вершины v и каждой длины блуждания T храним $h(v, T)$
- Заполняем его последовательно для возрастающих T

Моменты достижения в вершину

- Как это реализуется?
- Храним двумерный массив: для каждой вершины v и каждой длины блуждания T храним $h(v, T)$
- Заполняем его последовательно для возрастающих T
- Для $T = 0$ заполняем массив по всем вершинам нулями

Моменты достижения в вершину

- Как это реализуется?
- Храним двумерный массив: для каждой вершины v и каждой длины блуждания T храним $h(v, T)$
- Заполняем его последовательно для возрастающих T
- Для $T = 0$ заполняем массив по всем вершинам нулями
- Для каждого следующего T вычисляем каждую ячейку из значений для $T - 1$

Моменты достижения в вершину

- Скольку времени это займет?

Моменты достижения в вершину

- Скольку времени это займет?
- Для каждой длины от 0 до T проходим по всем вершинам

Моменты достижения в вершину

- Скольку времени это займет?
- Для каждой длины от 0 до T проходим по всем вершинам
- Для каждой вершины нужно столько операций, сколько есть соседей у вершины

Моменты достижения в вершину

- Скольку времени это займет?
- Для каждой длины от 0 до T проходим по всем вершинам
- Для каждой вершины нужно столько операций, сколько есть соседей у вершины
- По сути число операций будет порядка суммы степеней вершин

Моменты достижения в вершину

- Скольку времени это займет?
- Для каждой длины от 0 до T проходим по всем вершинам
- Для каждой вершины нужно столько операций, сколько есть соседей у вершины
- По сути число операций будет порядка суммы степеней вершин
- Это удвоенное число ребер!

Моменты достижения в вершину

- Скольку времени это займет?
- Для каждой длины от 0 до T проходим по всем вершинам
- Для каждой вершины нужно столько операций, сколько есть соседей у вершины
- По сути число операций будет порядка суммы степеней вершин
- Это удвоенное число ребер!
- Нужно порядка $T \cdot E$ операций

Время перемещения

- У нас был и симметричный вариант меры близости: усеченное время перемещения

Время перемещения

- У нас был и симметричный вариант меры близости: усеченное время перемещения
- Сумма моментов достижения в обе стороны

Время перемещения

- У нас был и симметричный вариант меры близости: усеченное время перемещения
- Сумма моментов достижения в обе стороны
- Как вычислять время перемещения из данной вершины во все остальные?

Время перемещения

- У нас был и симметричный вариант меры близости: усеченное время перемещения
- Сумма моментов достижения в обе стороны
- Как вычислять время перемещения из данной вершины во все остальные?
- Просто применить два описанных метода и сложить!

Случайные блуждания

Задачи на графах

Случайные блуждания

Случайные блуждания в приложениях

Моменты достижения из вершины

Моменты достижения в вершину

Блуждания длины 3

Число общих соседей

- Напомним, что у нас уже был простой способ находить ближайшие вершины

Число общих соседей

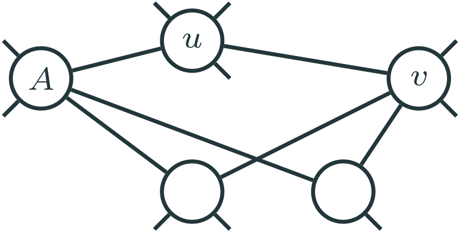
- Напомним, что у нас уже был простой способ находить ближайшие вершины
- Просто смотрели на число общих соседей

Число общих соседей

- Напомним, что у нас уже был простой способ находить ближайшие вершины
- Просто смотрели на число общих соседей
- Оказывается, этот подход близок к тому, что получается при блужданиях длины 3

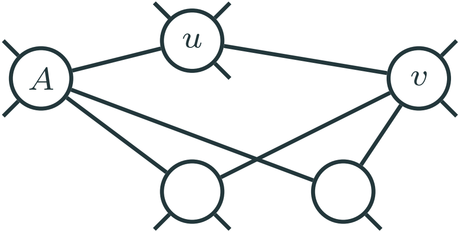
Блуждания из вершины

- Посмотрим на 3-усеченный момент из вершины A в вершину v



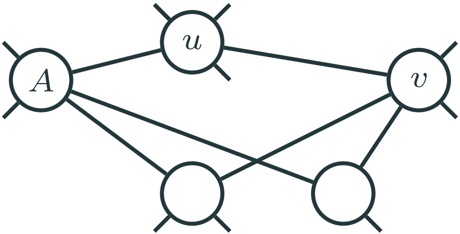
Блуждания из вершины

- Посмотрим на 3-усеченный момент из вершины A в вершину v
- Нас интересуют только не соединенные вершины



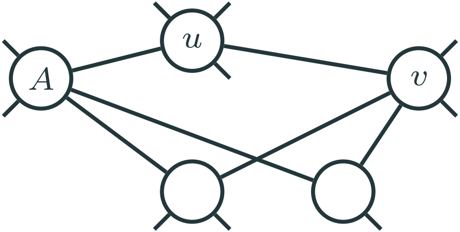
Блуждания из вершины

- Посмотрим на 3-усеченный момент из вершины A в вершину v
- Нас интересуют только не соединенные вершины
- Так что нет шансов дойти за один ход



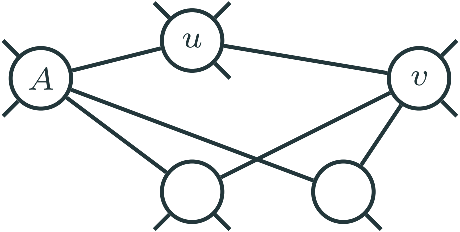
Блуждания из вершины

- Каковы шансы дойти за два хода?



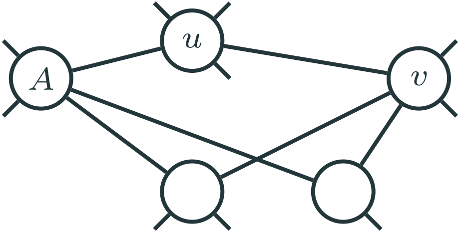
Блуждания из вершины

- Каковы шансы дойти за два хода?
- Нужно два события: на первом шаге перейти в соседа v и на втором шаге перейти в v



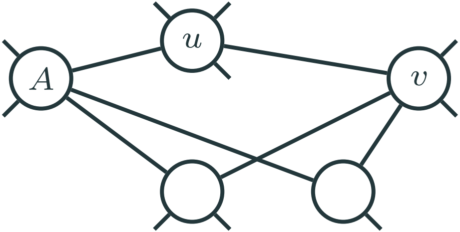
Блуждания из вершины

- Каковы шансы дойти за два хода?
- Нужно два события: на первом шаге перейти в соседа v и на втором шаге перейти в v
- Первое событие означает, что мы перешли в общего соседа A и v



Блуждания из вершины

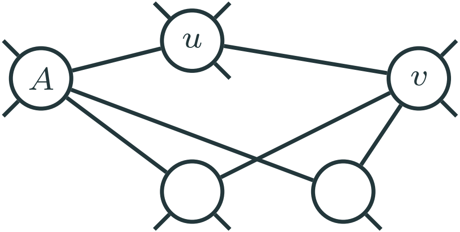
- Если мы попали в вершину u, то второе событие происходит с вероятностью $1/d(u)$



Блуждания из вершины

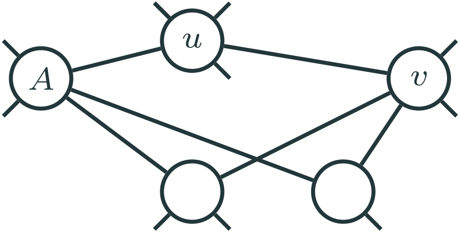
- Если мы попали в вершину u, то второе событие происходит с вероятностью $1/d(u)$
- Суммарная вероятность получается равной

$$p = \sum_{u \in N(A,v)} \frac{1}{d(A)} \cdot \frac{1}{d(u)} = \frac{1}{d(A)} \cdot \sum_{u \in N(A,v)} \frac{1}{d(u)}$$



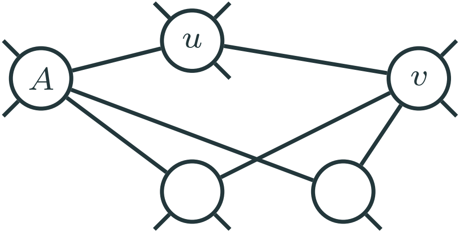
Блуждания из вершины

- С вероятностью $1 - p$ длина пути равна 3



Блуждания из вершины

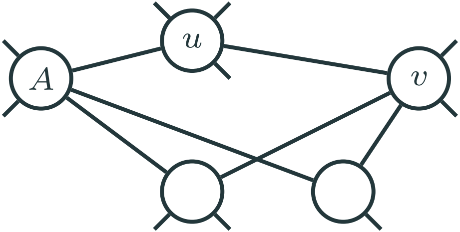
- С вероятностью $1 - p$ длина пути равна 3
- Средняя длина пути тем короче, чем больше p



Блуждания из вершины

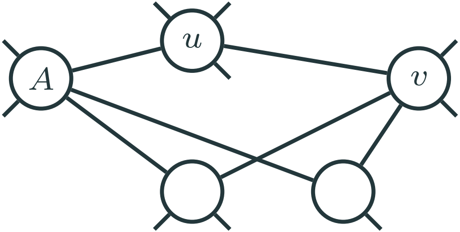
- По сути p меряет число общих соседей A и v, но каждый сосед учитывается с весом равным 1, деленым на его степень:

$$p = \frac{1}{d(A)} \cdot \sum_{u \in N(A,v)} \frac{1}{d(u)}$$



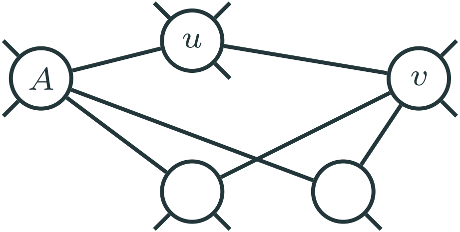
Блуждания из вершины

- Чем меньше степень соседа, тем больше его вклад в сумму



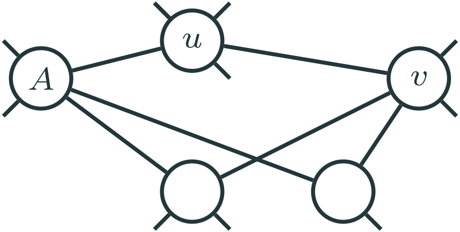
Блуждания из вершины

- Чем меньше степень соседа, тем больше его вклад в сумму
- Это достаточно естественно с точки зрения нашей задачи



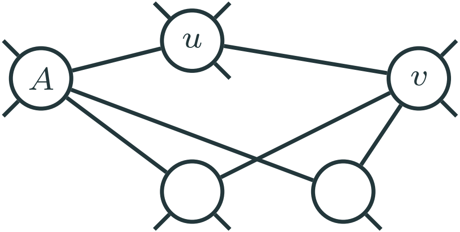
Блуждания из вершины

- Если смотреть на блуждания в данную вершину, то будет похожий результат
- $p = \frac{1}{d(v)} \cdot \sum_{u \in N(A,v)} \frac{1}{d(u)}$



Блуждания из вершины

- Если смотреть на блуждания в данную вершину, то будет похожий результат
- $p = \frac{1}{d(v)} \cdot \sum_{u \in N(A,v)} \frac{1}{d(u)}$
- Мы делим на степень вершины-кандидата, грубо говоря, измеряем долю общих соседей A и v среди соседей v



Что мы получили

- Итак, у нас есть 4 способа искать вершины, с которыми стоит соединить данную

Что мы получили

- Итак, у нас есть 4 способа искать вершины, с которыми стоит соединить данную
    - Простой способ: число общих соседей

Что мы получили

- Итак, у нас есть 4 способа искать вершины, с которыми стоит соединить данную
    - Простой способ: число общих соседей
    - Три способа, основанных на случайных блужданиях

Что мы получили

- Итак, у нас есть 4 способа искать вершины, с которыми стоит соединить данную
  - Простой способ: число общих соседей
  - Три способа, основанных на случайных блужданиях
- Попробуем посмотреть, что из этого лучше работает на практике