

Сетевой анализ

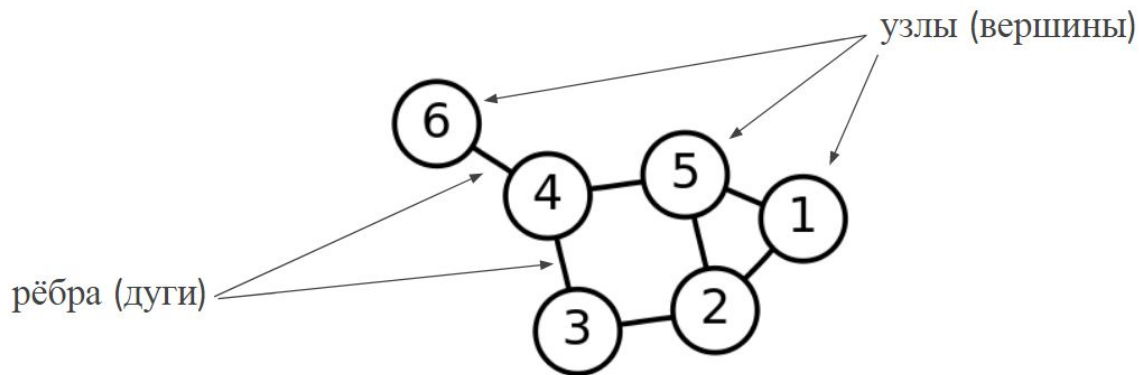
Презентация подготовлена с использованием материалов, любезно предоставленных А.Б. Хазовой

Что такое граф?

Граф, или сеть – это модель, состоящая из узлов и связей между ними.

Узлы принято называть **вершинами**, а связи – **ребрами**.

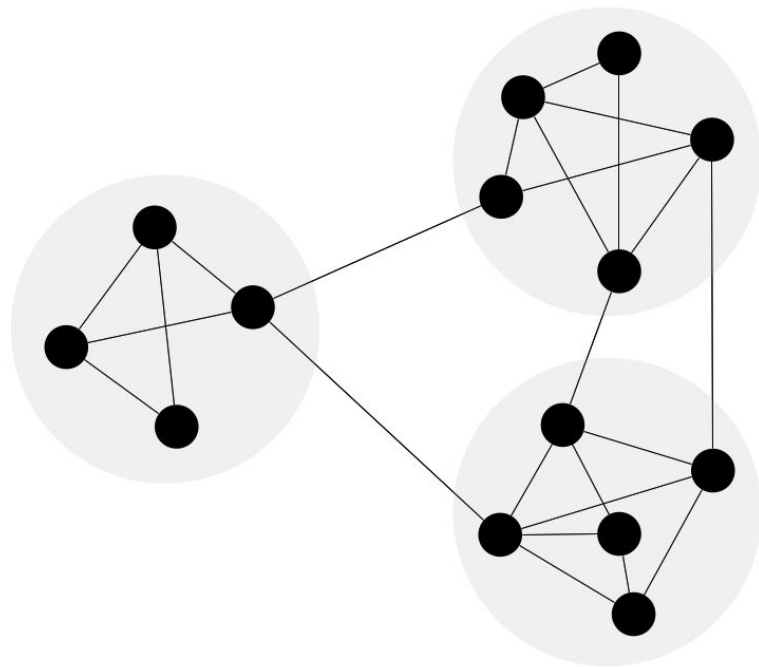
По-английски nodes (vertices) и edges.



Что такое граф?

Узлы в графах могут группироваться в сообщества.

Сообщество – это плотный подграф, где все (или почти все) узлы связаны между собой.



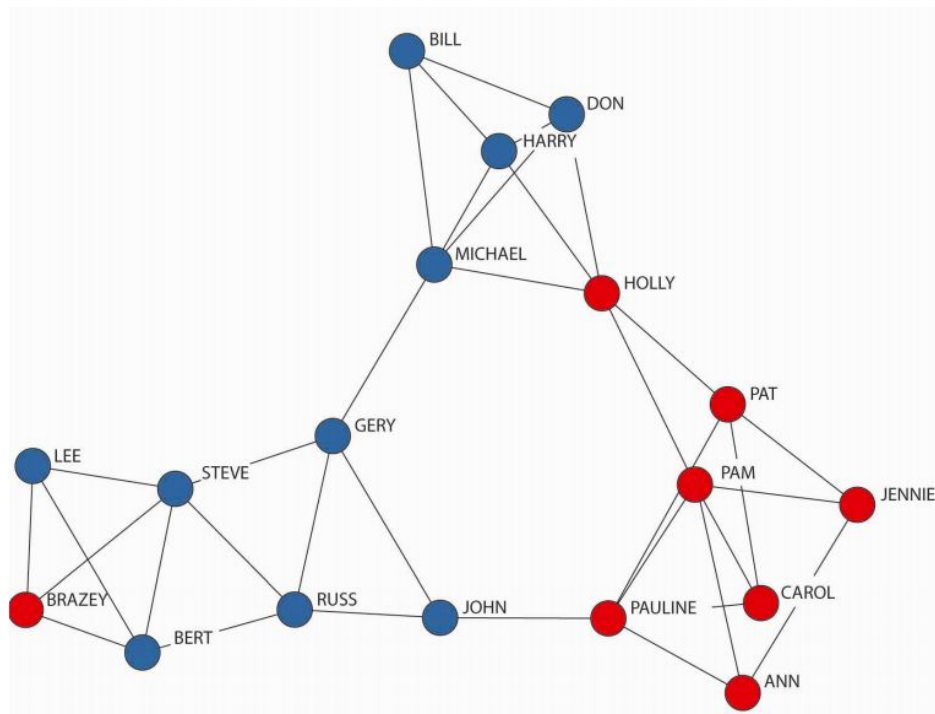
Типы графов

Графы бывают:

- **ориентированные и неориентированные**

Наличие/отсутствие указания на направление связи

Неориентированный граф



Типы графов

Графы бывают:

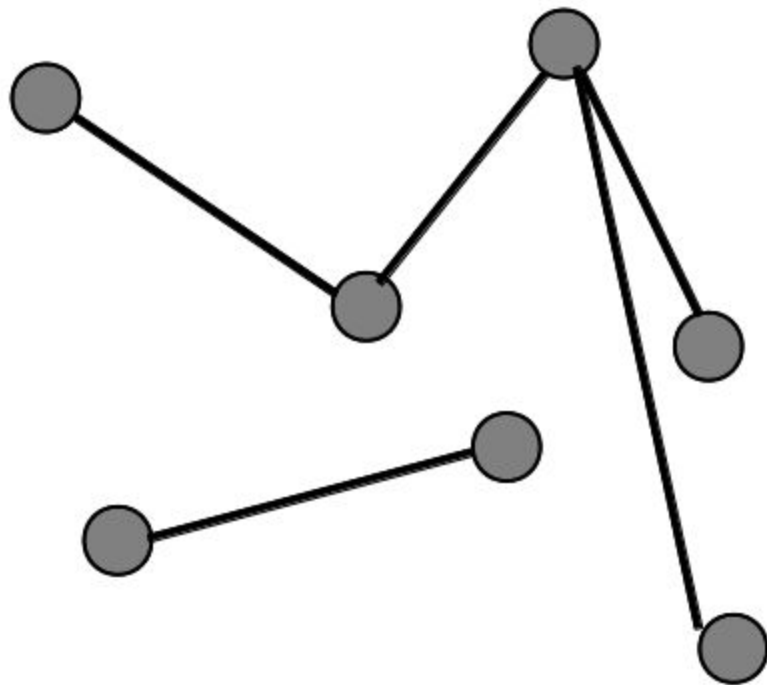
- **ориентированные и неориентированные**

Наличие/отсутствие указания на направление связи

- **связные и несвязные**

Все узлы связаны vs есть узлы, которые оторваны от основного графа

Несвязный граф



Это тоже граф



Типы графов

Графы бывают:

- **ориентированные и неориентированные**

Наличие/отсутствие указания на направление связи

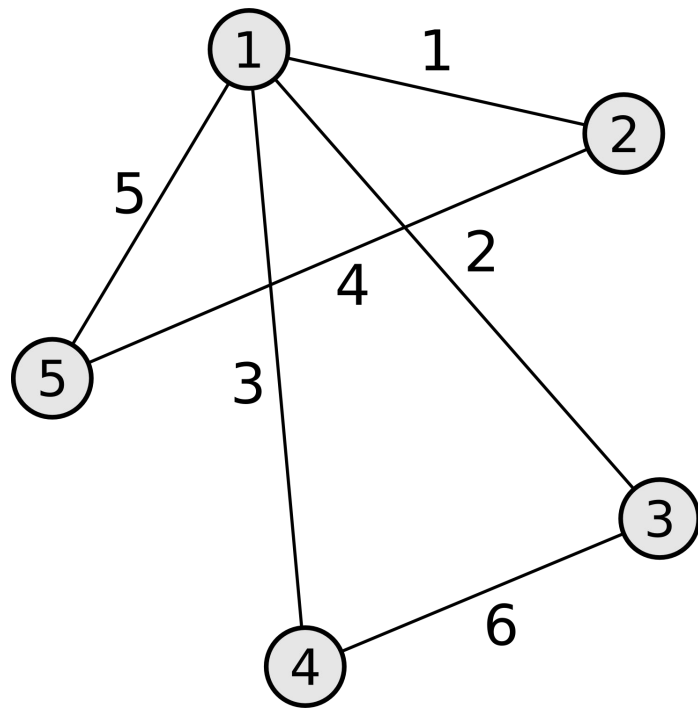
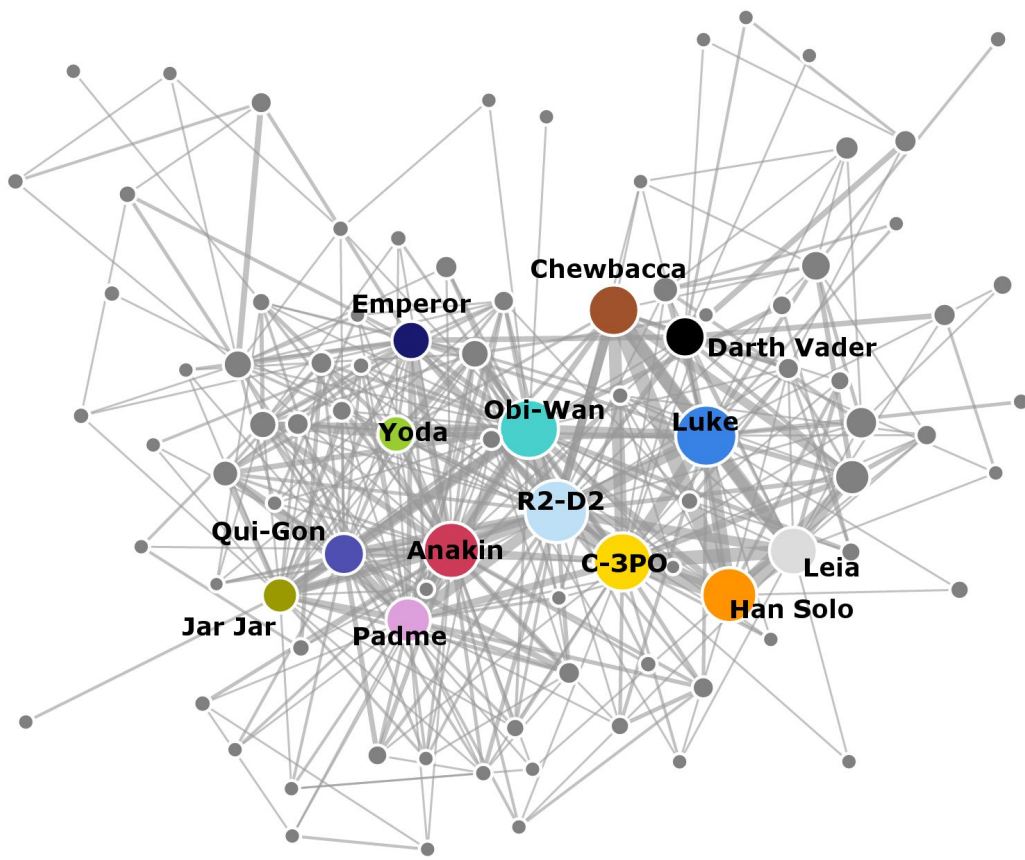
- **связные и несвязные**

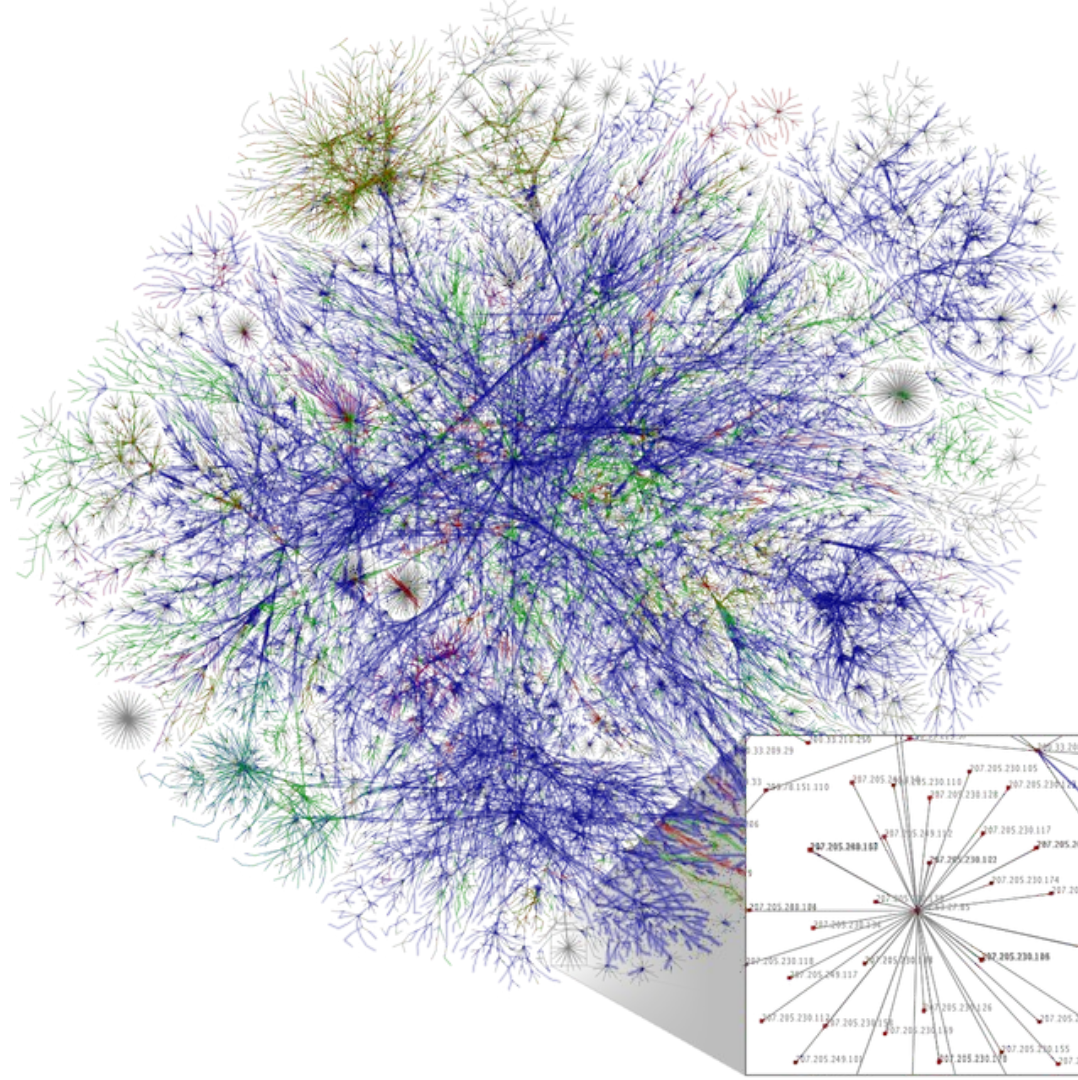
Все узлы связаны vs есть узлы, которые оторваны от основного графа

- **взвешенные и невзвешенные**

Связи имеют разный вес или равноправны

Взвешенные графы





Метрики

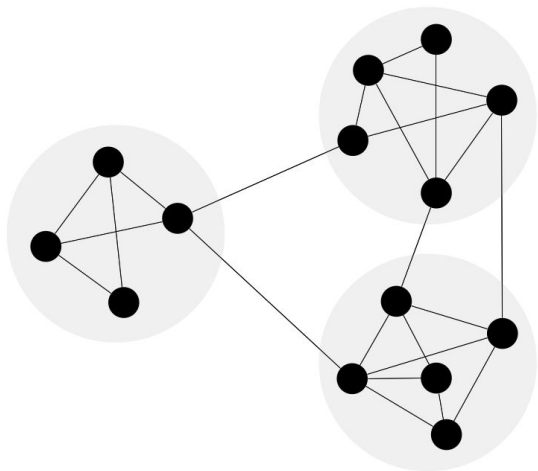
Метрика – это результат измерений, проведенных определенным способом.

Представьте, что вы выбираете **материал для реферата** по “Философским крохам” Кьеркегора. У вас есть **оригинальный текст в 50 страниц**, **современная книга** “Cumulative Index to Kierkegaard's Writings: The Works of Søren Kierkegaard” Н.Ж. Хонг и **статья** “Идея спасения в псевдонимных произведениях С. Керкегора. Очерк первый. Лестница Йоханнеса Климакуса” Д.А. Лунгиной **в 100 страниц**.

Если метрикой для вас является **количество страниц**, то вы выберете оригинальный текст, а если **простота чтения** – то статью.

Метрики

- **Степень, или мощность узла (degree)** – это количество его связей.
- **Взвешенная степень (weighed degree)** – это количество связей узла, разделенное на общее количество связей в графе.

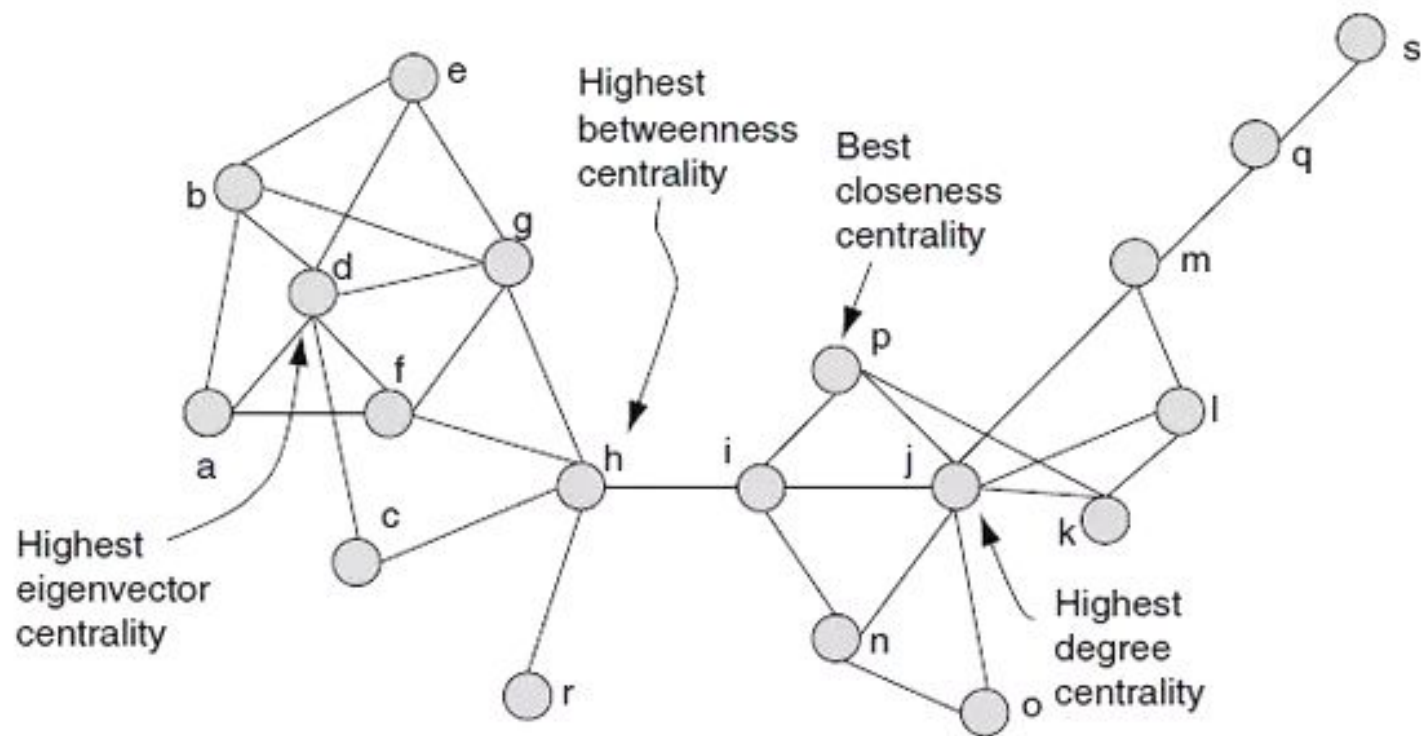


Метрики

Важность узла можно определять разными способами:

- **degree centrality:** у кого больше связей, тот и важнее
- **closeness centrality:** чем центральнее узел (т.е. чем короче путь от него до всех остальных узлов), тем он важнее
- **betweenness centrality:** количество кратчайших путей, проходящих через узел
- **eigencentrality:** чем больше связей у соседних узлов, тем узел важнее, иначе говоря, “чем больше друзей у твоих друзей, тем ты важнее”

Метрики графа



Метрики

- **Коэффициент ассортативности (assortativity coefficient)** определяет, с кем связаны "важные" узлы: если с другими "важными" узлами, то значение коэффициента высокое, а если нет – низкое.

Метрики

- **Коэффициент ассортативности (assortativity coefficient)** определяет, с кем связаны "важные" узлы: если с другими "важными" узлами, то значение коэффициента высокое, а если нет – низкое.
- **Коэффициент кластеризации (clustering coefficient)** – степень взаимодействия между собой ближайших соседей узла, т.е. вероятность того, что ближайшие соседи узла будут связаны не только с ним, но и между собой.

Метрики

- **Плотность графа (density)** – отношение числа ребер к максимально возможному. В сообществах высокий коэффициент кластеризации и высокая плотность.

Метрики

- **Плотность графа (density)** – отношение числа ребер к максимально возможному. В сообществах высокий коэффициент кластеризации и высокая плотность.
- **Модулярность (modularity)** показывает, насколько при заданном разбиении графа на группы плотность связей внутри группы больше плотности связей между группами. С помощью этой метрики граф разбивается на сообщества.

Метрики

- **Плотность графа (density)** – отношение числа ребер к максимально возможному. В сообществах высокий коэффициент кластеризации и высокая плотность.
- **Модулярность (modularity)** показывает, насколько при заданном разбиении графа на группы плотность связей внутри группы больше плотности связей между группами. С помощью этой метрики граф разбивается на сообщества.

Сети коллокаций

Сеть коллокаций – это визуализация коллокационных отношений некоторого выбранного слов с его коллокатами.

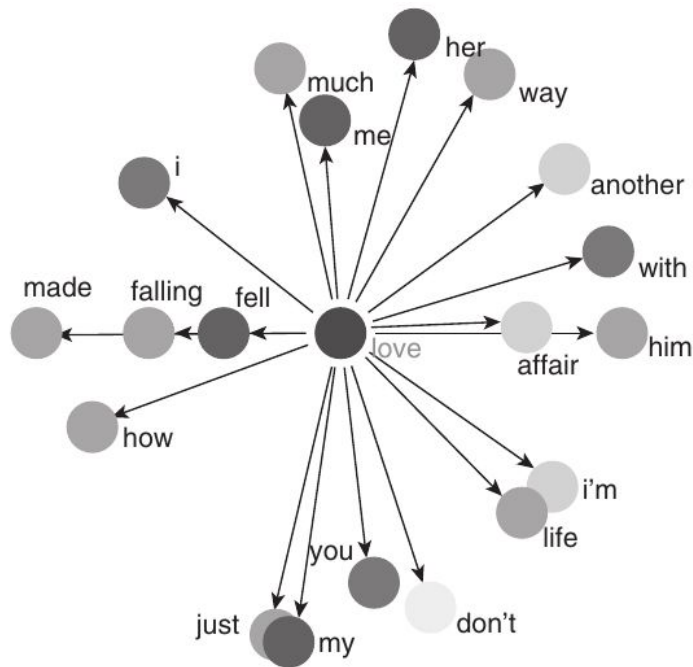
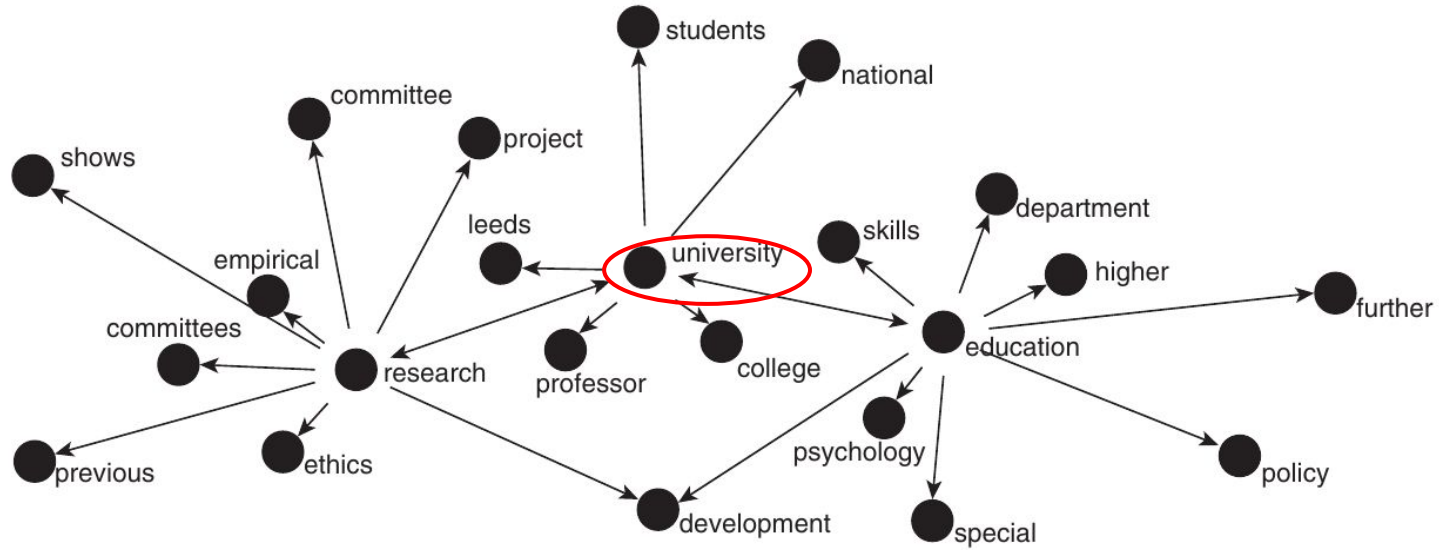


Figure 3.2 Collocation graph: 'love' in BE06 (10a – log Dice (7), L3–R3, C5–NC5)

‘University’



Как это может нам пригодиться?

Графы позволяют проследить, какие средства выражения чаще встречаются в рамках одной переводной единицы, то есть чаще выступают в качестве переводных эквивалентов друг друга.

На семантической карте это проследить сложнее, т.к. эта информация рассредоточена по нескольким картинкам. Сколько языков в выборке, столько и картинок.

Как это устроено?

Узлы – ?

Как это устроено?

Узлы – средства выражения из разных языков, например, должен (ru), il faut (fr), tener que (es) и т.д.

Как это устроено?

Узлы – средства выражения из разных языков, например, должен (ru), il faut (fr), tener que (es) и т.д.

Ребра – ?

Как это устроено?

Узлы – средства выражения из разных языков, например, должен (ru), il faut (fr), tener que (es) и т.д.

Ребра – указание на совместную встречаемость в переводной единице.

Вес – частота совместной встречаемости

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

I

I

C

I

E

I

Как э

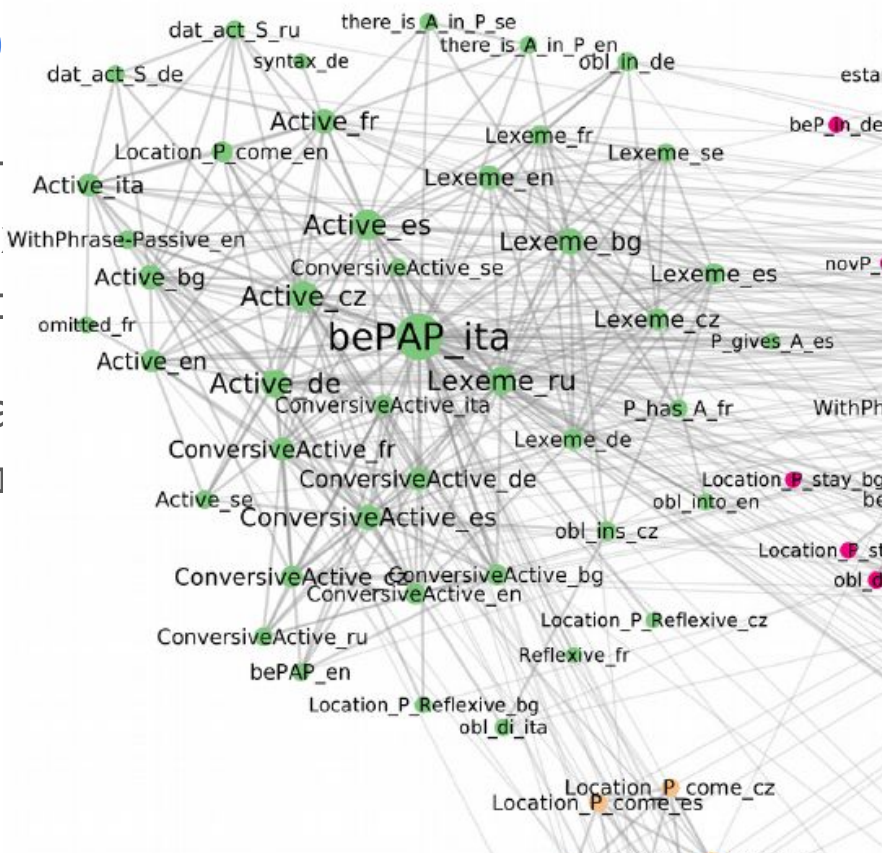
Графы г
в рамка
перевод

На семью
распред
столько

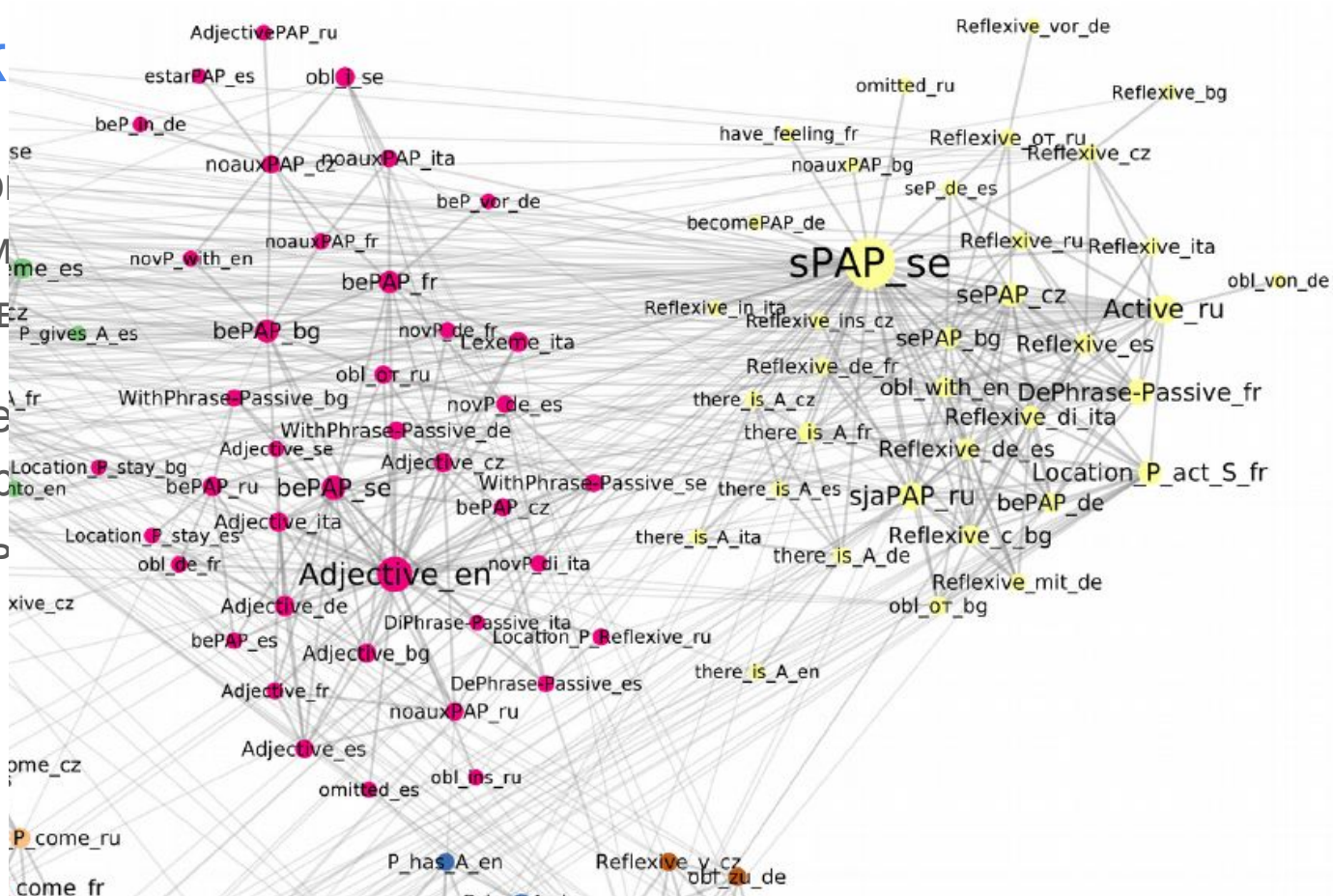
я?

выражения чаще встречаются
аще выступают в качестве

ее, т.к. эта информация
только языков в выборке,



Граф
в рам
перев
На се
расср
столь

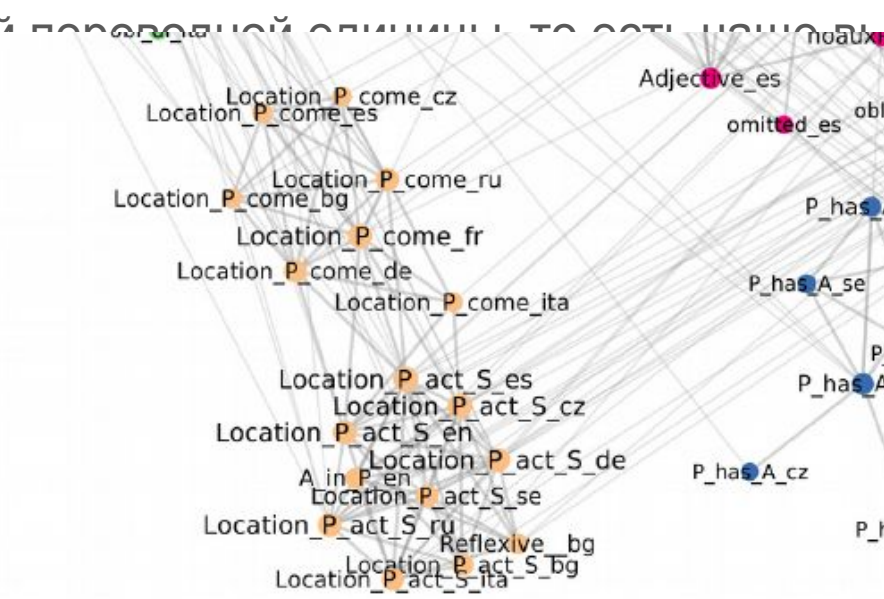


речаются
честве
ция
ке,

Как это может нам пригодиться?

Графы позволяют проследить, какие средства выражения чаще встречаются в рамках одной переводной единицы, то есть, часто выступают в качестве переводных.

На семанти-
рассредоточе-
столько и к



эта информация
языков в выборке,