Факультет экономических наук

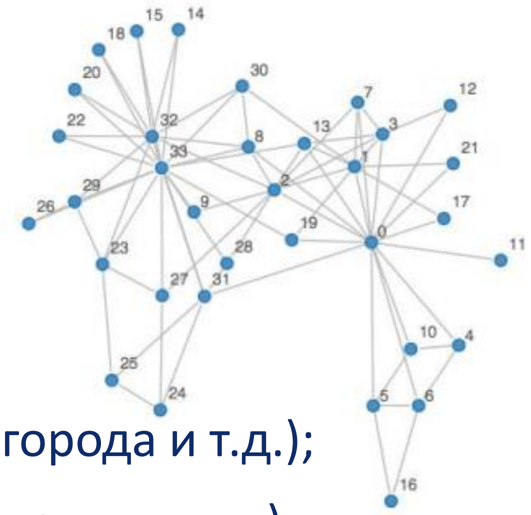
Департамент математики

Москва

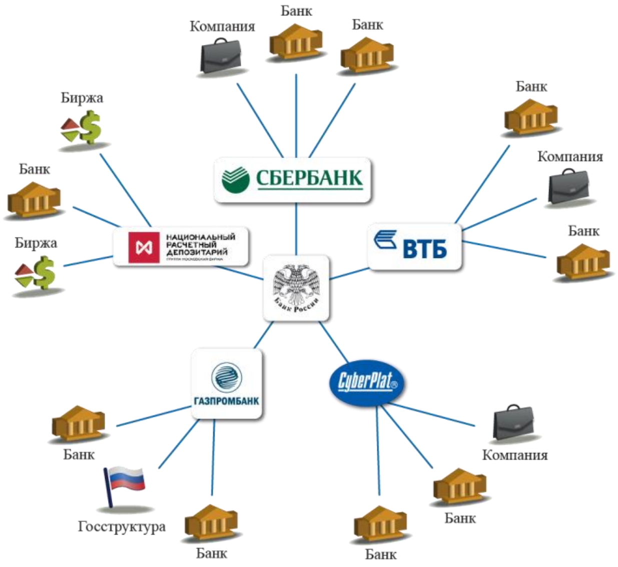
Графы и сети

Егорова Людмила

ЧТО ТАКОЕ СЕТЬ?

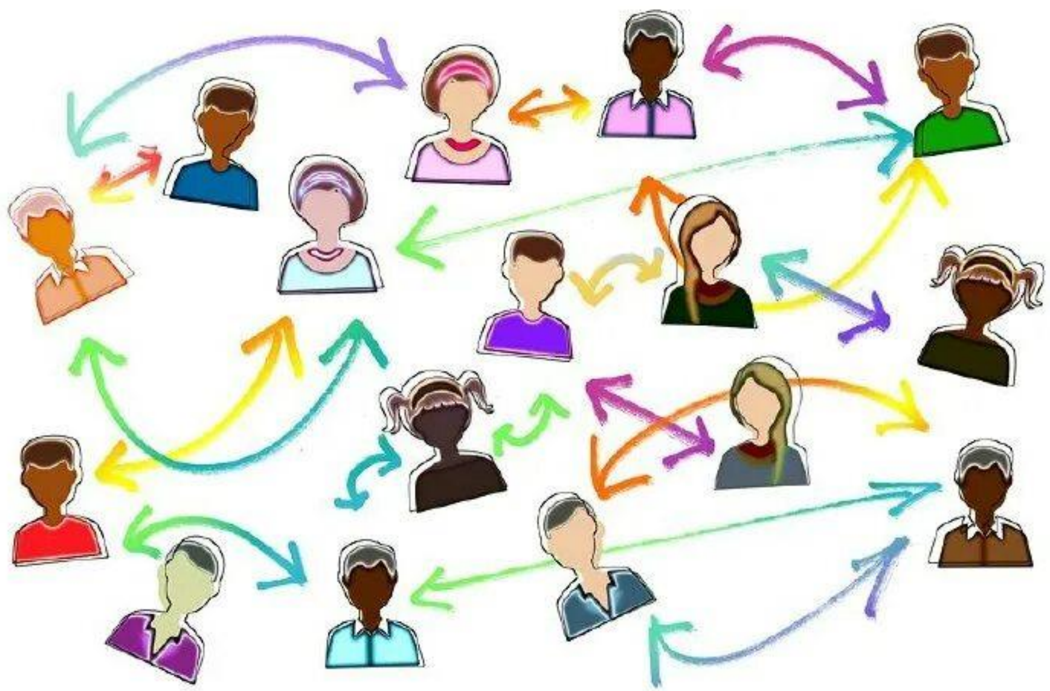


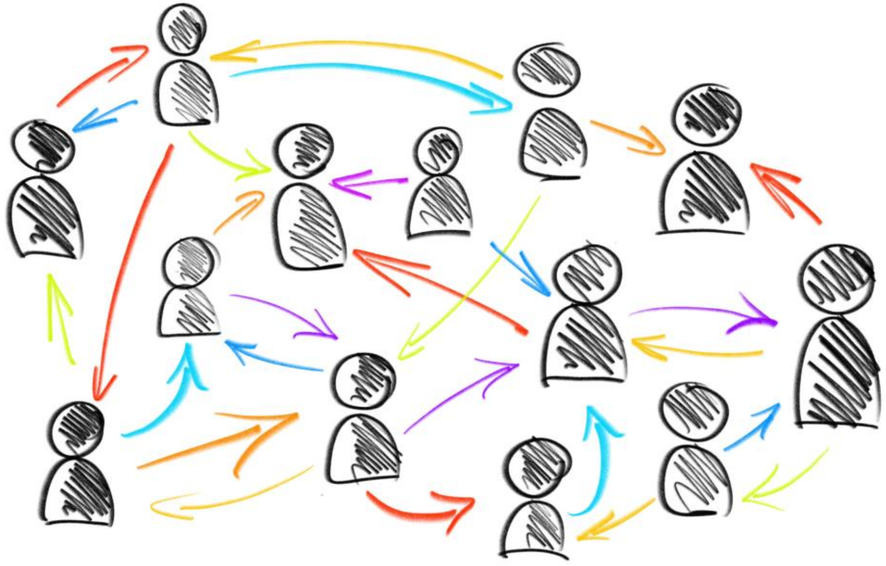
- Сеть представляет собой граф $G = G(V, E)$, в котором
- V – множество вершин/узлов (объекты, люди, финансовые институты, страны, города и т.д.);
- E – множество ребер. (взаимосвязи, отношения, дороги, потоки, телефонные звонки и т.д.).



Виды сетей

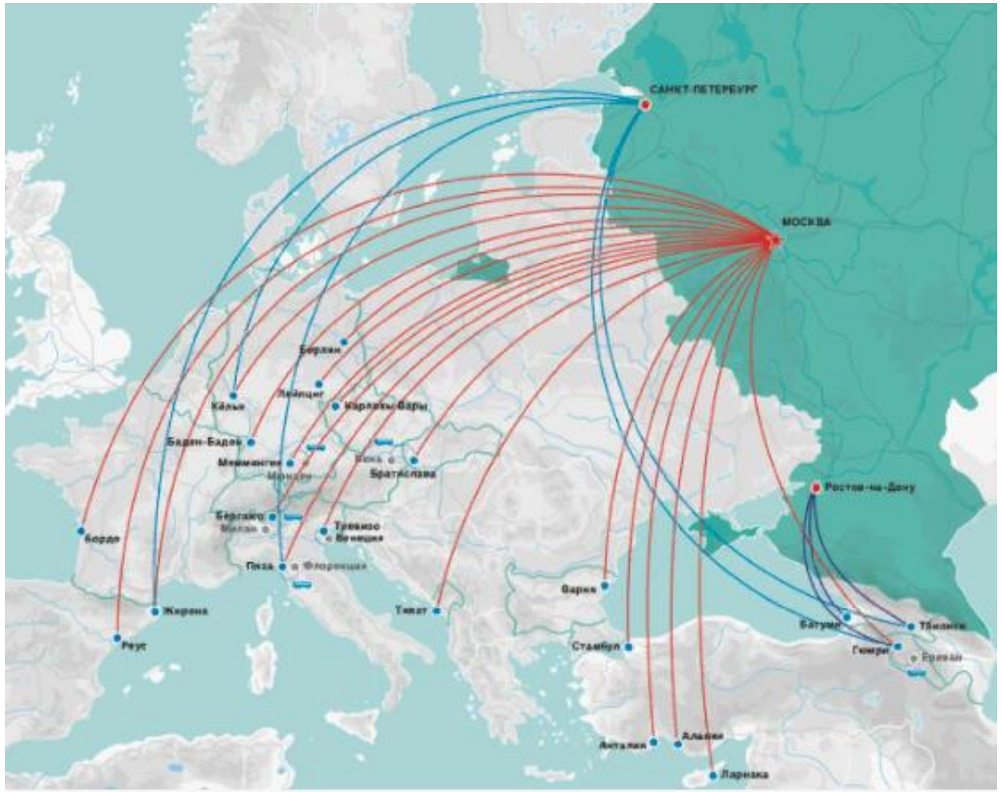
- Ориентированные и неориентированные графы

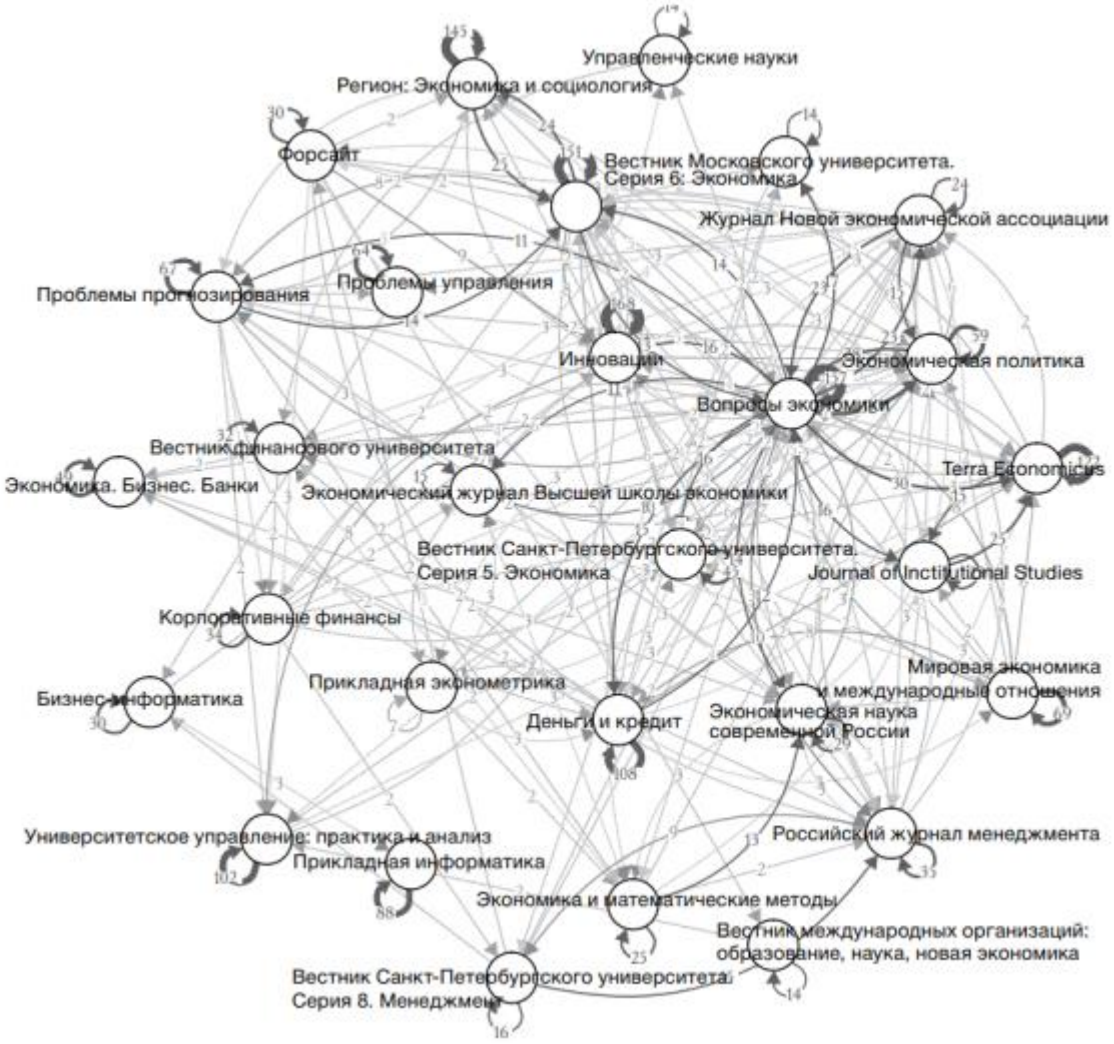




Виды сетей

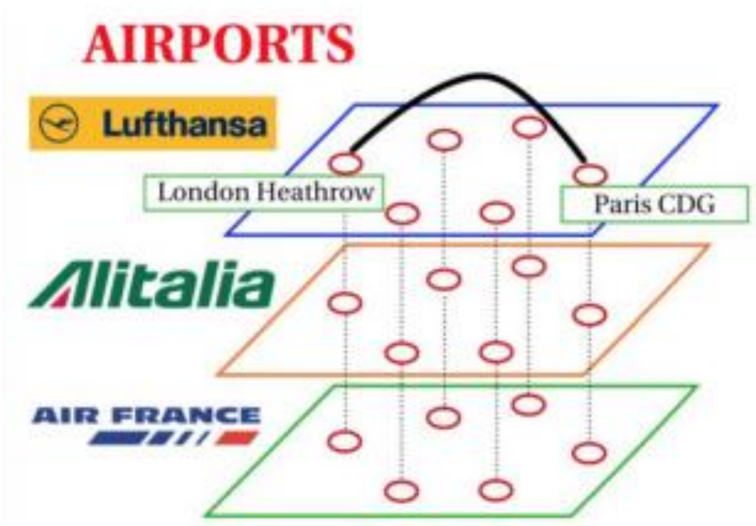
• Взвешенные и невзвешенные графы

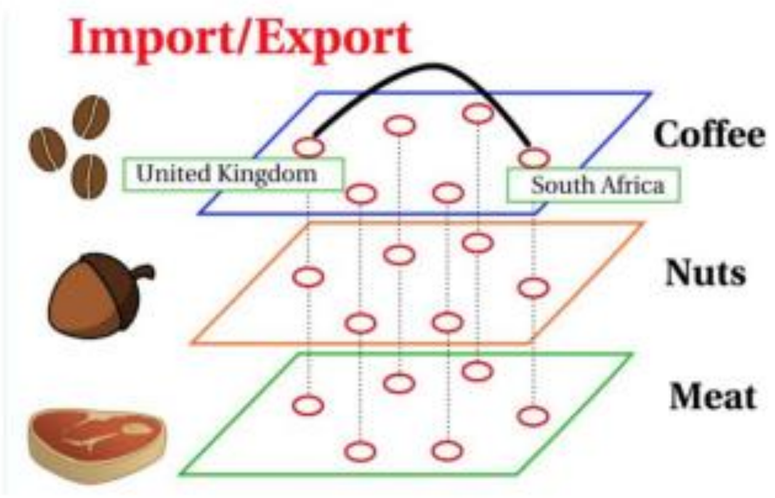




Виды сетей

- Многослойные сети





ПРЕДСТАВЛЕНИЕ СЕТЕВЫХ СТРУКТУР

- Списочное представление (вершин и ребер) сети

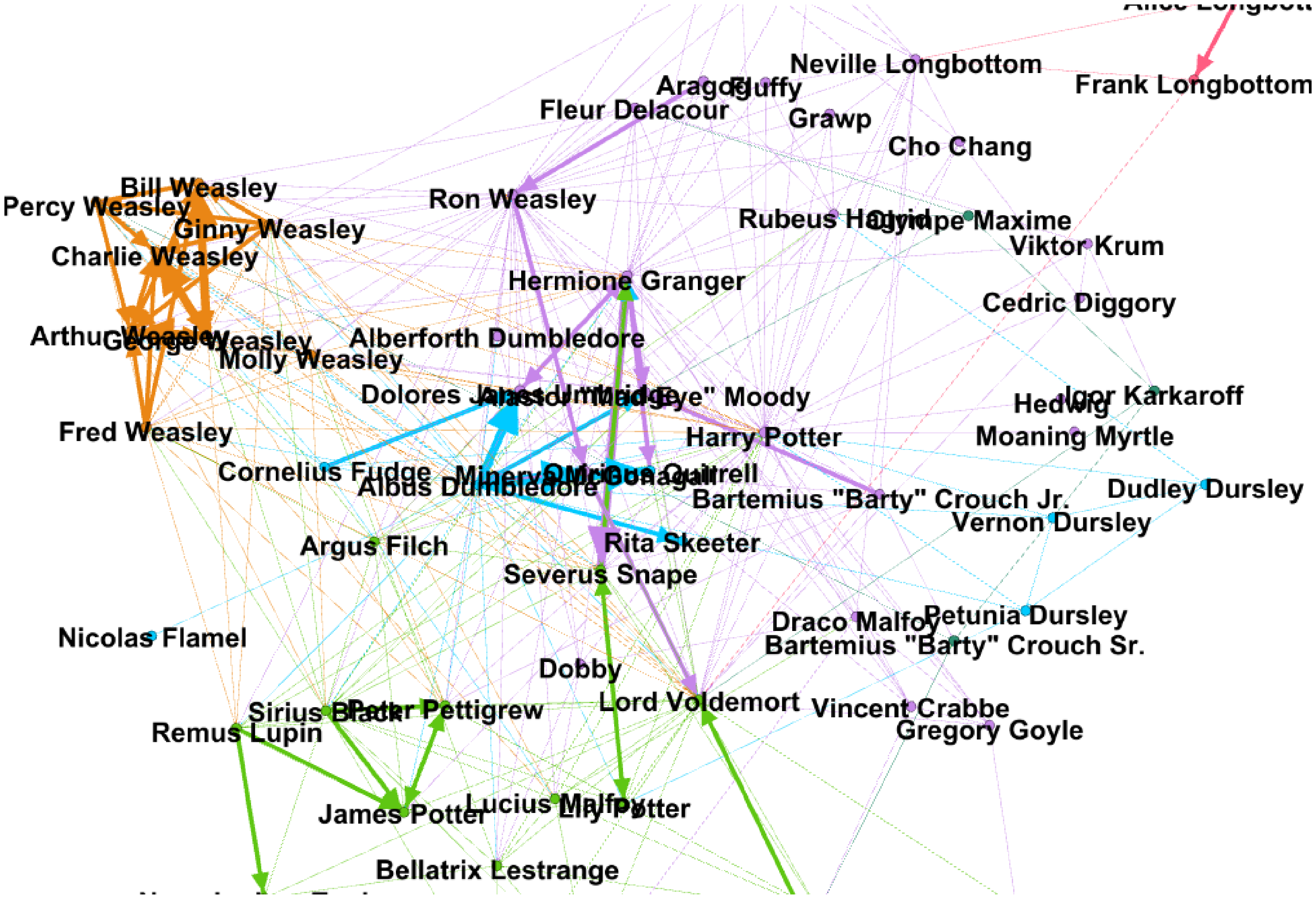
id	name	bio
0	Regulus Arcturus Black	Brother of Sirius. Used to be a Death Eater but defected.
1	Sirius Black	Best friend of James Potter and godfather of Harry.
2	Lavender Brown	Killed by a werewolf. She was a gryffindor student who dated Ron.
3	Cho Chang	Ravenclaw student who dated Cedric Diggory and Harry Potter.
4	Vincent Crabbe Sr.	Father of Crabbe and death-eater who escaped Azkaban.
5	Vincent Crabbe	Slytherin student who was best friends with Goyle and followed Draco.
6	Bartemius "Barty" Crouch Sr.	Head of the department of Internation Magical Cooperation. Killed by his son.
7	Bartemius "Barty" Crouch Jr.	Death Eater who impersonated Alastor Moody.
8	Fleur Delacour	Participated in the Triwizard tournament and married Bill Weasley.
9	Cedric Diggory	Participated in the Triwizard tournament and got killed by Voldemort.

source	target	type
0	1	-
0	25	-
0	45	-
1	0	-
1	11	+
1	21	+
1	25	-
1	31	+
1	33	-
1	34	-
1	36	+

- В виде графа

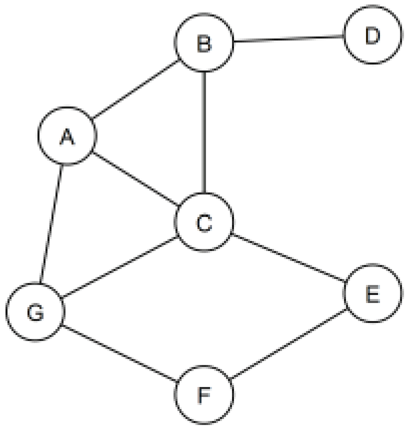
ПРЕДСТАВЛЕНИЕ СЕТЕВЫХ СТРУКТУР

- В виде графа



ПРЕДСТАВЛЕНИЕ СЕТЕВЫХ СТРУКТУР

- В виде матрицы смежности.
- A - матрица смежности графа G, где $a_{ij} = 1$, если между вершинами i и j есть связь, и 0 иначе.
- В общем случае ориентированного графа a_{ij} может быть не равно a_{ji}.
- В случае взвешенного графа будем также рассматривать матрицу весов W, в которой значение w_{ij} для связанных вершин i и j есть вес ребра ij, и $w_{ij} = 0$, если нет ребра ij, т.е. нет связи между I и j.



$$
\begin{array}{c c}
 & \begin{array}{ccccccc} A & B & C & D & E & F & G \end{array} \\
\begin{array}{c} A \\ B \\ C \\ D \\ E \\ F \\ G \end{array} &
\begin{pmatrix}
0 & 1 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 & 1 & 0
\end{pmatrix}
\end{array}
$$

ЧТО ТАКОЕ СЕТЬ?

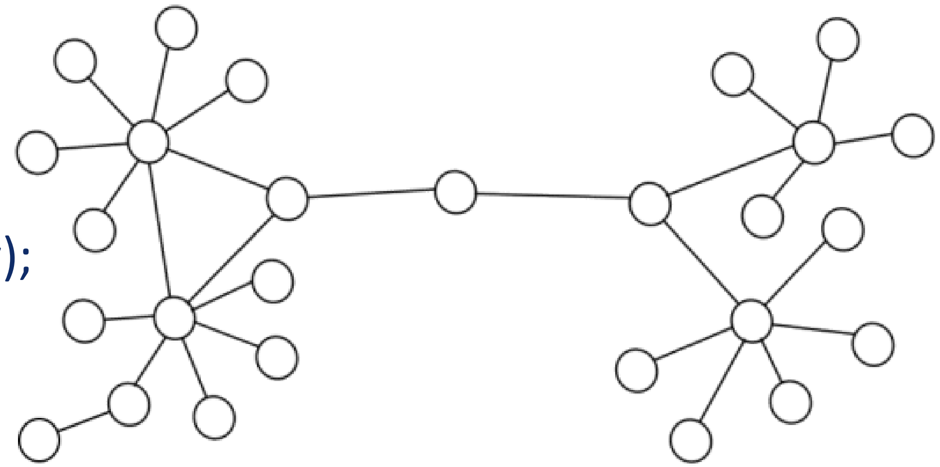
- Анализ сетевых моделей может быть важен:

- для выявления важных, ключевых узлов и компонентов сети, обеспечивающих устойчивость системы (сети) в целом по отношению к внешним и внутренним угрозам;
- в задаче анализа динамики распространения чего-либо по сети (информации или рекламы для маркетинга и таргетинга рекламной кампании, заражения и распространения болезни для анализа развития эпидемий, каскада экономических коллапсов и банкротств в банковской или предпринимательской сферах и т.д.);
- для выявления сообществ (сильно связанных между собой множеств вершин) в сети для декомпозиции сети на меньшие фрагменты;
- для описания динамики развития сетей с точки зрения предсказания будущих связей и прогноза сети.

Влияние в сети

- Центральность вершин в сети - это вектор, сопоставляющий каждой вершине сети некоторое число (индекс) так, чтобы значения индекса обеспечивали некоторое ранжирование (слабый порядок), на основе которого определяются наиболее важные элементы сети.

- Существуют разные подходы для расчета индексов центральности в сети. Каждый индекс имеет свою интерпретацию и по-своему оценивает важность вершины в сети. Наиболее распространенные классические индексы:
- степенная центральность (degree centrality);
- центральность по близости (closeness centrality);
- центральность по посредничеству (betweenness centrality);
- центральность по собственному вектору (Eigenvector centrality);
- центральность PageRank.



Степенная центральность

- исходит из идеи о том, что наиболее важной (центральной) является вершина, имеющая наибольшее число связей с другими вершинами.

- В зависимости от вида графа сети существуют различные формулы расчета степенной центральности:

для невзвешенного неориентированного графа

$$C_i^{deg} = \sum_{j=1}^{n} a_{ij} = \sum_{j=1}^{n} a_{ji};$$

для невзвешенного ориентированного графа

$$C_i^{out-deg} = \sum_{j=1}^{n} a_{ij},$$

$$C_i^{in-deg} = \sum_{j=1}^{n} a_{ji};$$

для взвешенного неориентированного графа

$$C_i^{deg} = \sum_{j=1}^{n} a_{ij} = \sum_{j=1}^{n} a_{ji},$$

$$C_i^{w\ deg} = \sum_{j=1}^{n} w_{ij} = \sum_{j=1}^{n} w_{ji};$$

для взвешенного ориентированного графа

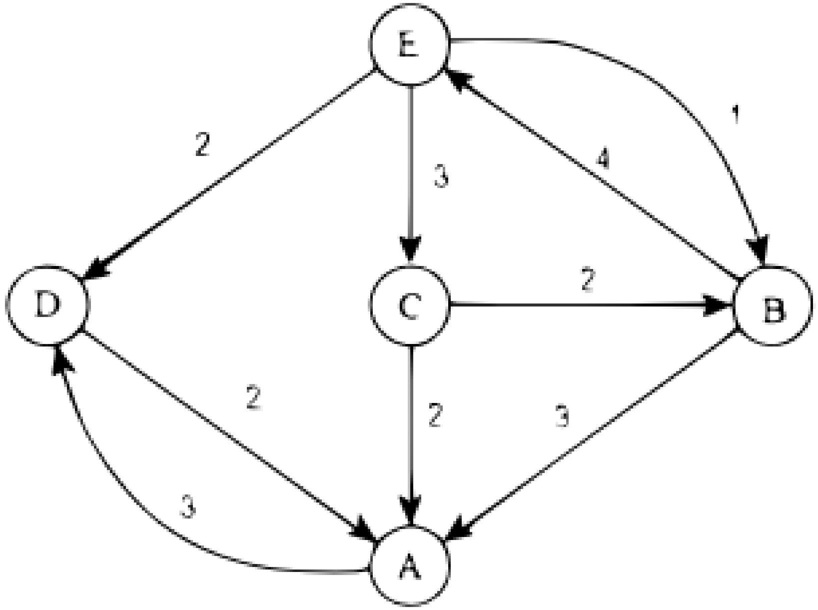
$$C_i^{in-deg} = \sum_{j=1}^{n} a_{ji},$$

$$C_i^{out-deg} = \sum_{j=1}^{n} a_{ij},$$

$$C_i^{w\ in-deg} = \sum_{j=1}^{n} w_{ji},$$

$$C_i^{w\ out-deg} = \sum_{j=1}^{n} w_{ij}.$$

Степенная центральность (пример)

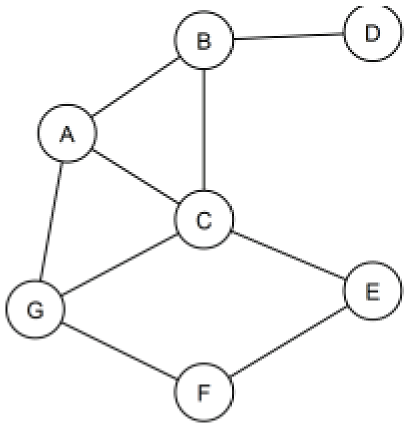


i	C_i^{in-deg}	$C_i^{out-deg}$	$C_i^{w\ in-deg}$	$C_i^{w\ out-deg}$
A	3	1	7	3
B	2	2	3	7
C	1	2	3	4
D	2	1	5	2
E	1	3	4	6

Центральность по близости

- исходит из идеи о том, что наиболее важной является вершина, находящаяся ближе всех к другим вершинам сети.
- Кратчайшим путем называется путь (цепь) между вершинами, имеющий наименьшую длину (длина пути - это количество ребер в этом пути).
- В неориентированном графе вершины, которые связаны напрямую (то есть являются концами одного ребра), называются смежными.

- Пример
- Найдите кратчайшие пути между:
- А и В;
- А и F;
- D и F.



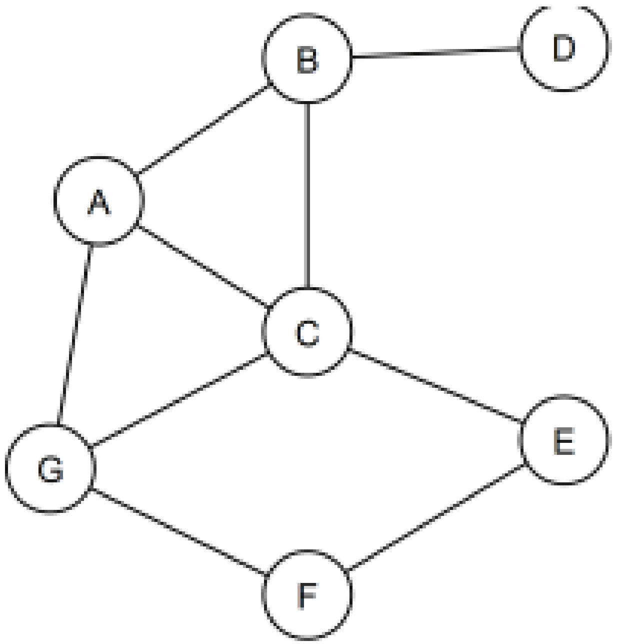
Центральность по близости

- Существуют две версии формулы центральности по близости (в общем случае они могут давать различающиеся результаты):

$$c_1(i) = \frac{1}{\sum_j d_{ij}}, \quad c_2(i) = \sum_j \frac{1}{d_{ij}}.$$

- Здесь d_{ij} обозначает длину кратчайшего пути от вершины i к вершине j.
- Часто для нормализации и сравнения сетей разных размеров значение центральности домножают на N-1, где N равно числу вершин в графе сети. Тогда индекс центральности по близости можно интерпретировать как среднее расстояние от выбранной вершины до всех остальных вершин сети.
- Центральность $c_2(i)$ называют еще гармонической центральностью.

Центральность по близости (пример)



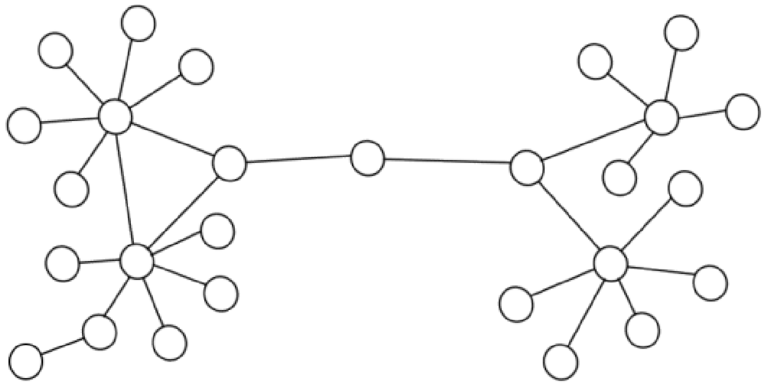
i	$c_1(i)$	$c_2(i)$
A	$\frac{1}{1+1+2+2+2+1} = \frac{1}{9}$	$\frac{1}{1}+\frac{1}{1}+\frac{1}{2}+\frac{1}{2}+\frac{1}{2}+\frac{1}{1}=\frac{9}{2}=4{,}5$
B	$\frac{1}{1+1+1+2+3+2} = \frac{1}{10}$	$\frac{1}{1}+\frac{1}{1}+\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\frac{1}{2}=\frac{13}{3}=4\frac{1}{3}$
C	$\frac{1}{1+1+2+1+2+1} = \frac{1}{8}$	$\frac{1}{1}+\frac{1}{1}+\frac{1}{2}+\frac{1}{1}+\frac{1}{2}+\frac{1}{1}=5$
D	$\frac{1}{2+1+2+3+4+3} = \frac{1}{15}$	$\frac{1}{2}+\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{3}=\frac{35}{12}=2\frac{11}{12}$
E	$\frac{1}{2+2+1+3+1+2} = \frac{1}{11}$	$\frac{1}{2}+\frac{1}{2}+\frac{1}{1}+\frac{1}{3}+\frac{1}{1}+\frac{1}{2}=\frac{23}{6}=3\frac{5}{6}$
F	$\frac{1}{2+3+2+4+1+1} = \frac{1}{13}$	$\frac{1}{2}+\frac{1}{3}+\frac{1}{2}+\frac{1}{4}+\frac{1}{1}+\frac{1}{1}=\frac{43}{12}=3\frac{7}{12}$
G	$\frac{1}{1+2+1+3+2+1} = \frac{1}{10}$	$\frac{1}{1}+\frac{1}{2}+\frac{1}{1}+\frac{1}{3}+\frac{1}{2}+\frac{1}{1}=\frac{13}{3}=4\frac{1}{3}$

Центральность по посредничеству

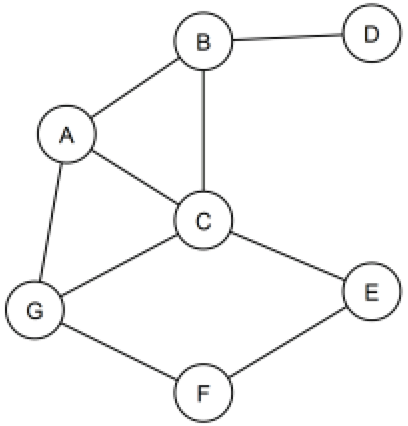
- считает наиболее важной (центральной) вершину, через которую проходит наибольшее число кратчайших путей между любыми другими парами вершин, т.е. важность (центральность) вершины здесь оценивается исходя из того, насколько часто вершина является посредником, соединяющим пары вершин сети:

$$b_i = \sum_{u,v \in V} \frac{\sigma_{uv}(i)}{\sigma_{uv}},$$

- где σ_{uv} - число кратчайших путей от вершины u к вершине v, $\sigma_{uv}(i)$ - число кратчайших путей от вершины u к вершине v, которые проходят через вершину i.



Центральность по посредничеству (пример)

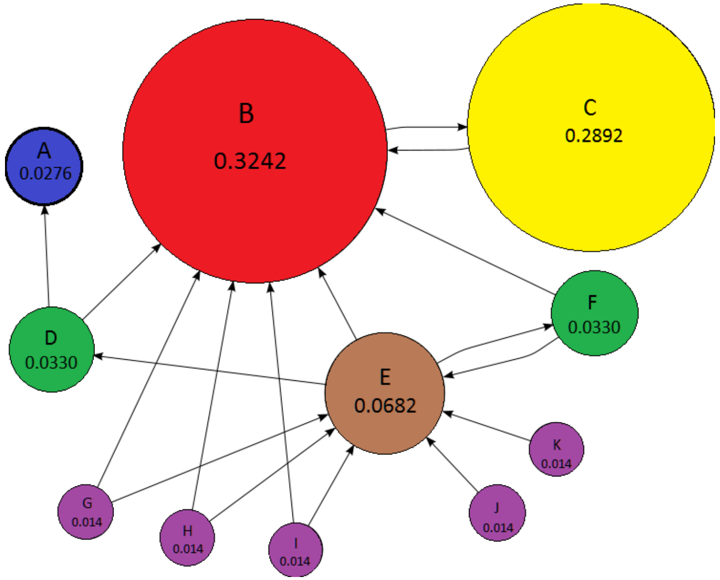


		A	B	C	D	E	F	G
Пара вершин u, v	Вершины, лежащие на кратчайшем пути от u до v	$\frac{\sigma_{uv}(A)}{\sigma_{uv}}$	$\frac{\sigma_{uv}(B)}{\sigma_{uv}}$	$\frac{\sigma_{uv}(C)}{\sigma_{uv}}$	$\frac{\sigma_{uv}(D)}{\sigma_{uv}}$	$\frac{\sigma_{uv}(E)}{\sigma_{uv}}$	$\frac{\sigma_{uv}(F)}{\sigma_{uv}}$	$\frac{\sigma_{uv}(G)}{\sigma_{uv}}$
A, D	B	-	1	-	-	-	-	-
A, E	C	-	-	1	-	-	-	-
A, F	G	-	-	-	-	-	-	1
B, G	A или C	1/2	-	1/2	-	-	-	-
B, E	C	-	-	1	-	-	-	-
B, F	A, G или C, G или C, E	1/3	-	2/3	-	1/3	-	2/3
C, D	B	-	1	-	-	-	-	-
C, F	E или G	-	-	-	-	1/2	-	1/2
D, G	A, B или B, C	1/2	1	1/2	-	-	-	-
D, E	B, C	-	1	1	-	-	-	-
D, F	B, A, G или B, C, G или B, C, E	1/3	1	2/3	-	1/3	-	2/3
E, G	C или F	-	-	1/2	-	-	1/2	-
Сумма		$1\frac{2}{3}$	5	$5\frac{5}{6}$	0	$1\frac{1}{6}$	1/2	$2\frac{5}{6}$

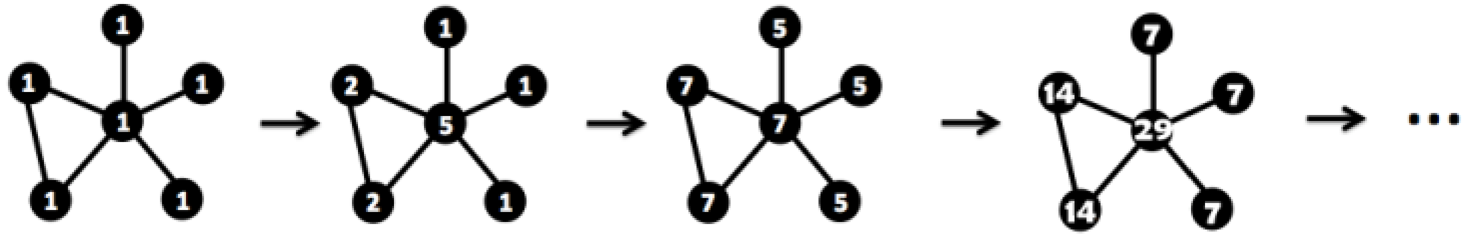
Центральность по собственному вектору

- основан на идее о том, что важность вершины C_i^{ev} зависит от важности ее соседей - чем важнее соседи, тем важнее сама вершина:

$$C_i^{ev} = \frac{1}{\lambda} \sum_{j:a_{ij} \neq 0} C_j^{ev} = \frac{1}{\lambda} \sum_{j=1}^{n} C_j^{ev} \cdot a_{ij},$$



- Для нахождения индекса центральности необходимо найти собственный вектор C^{ev} матрицы смежности А, соответствующий максимальному собственному значению А.



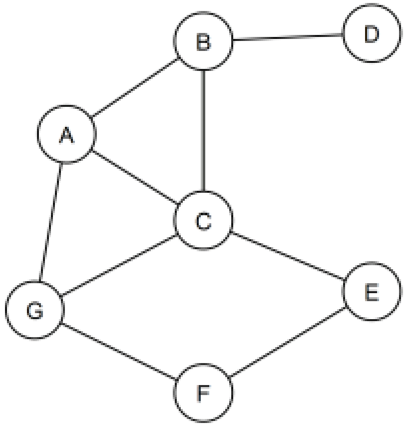
Центральность PageRank

- Эта центральность предложена для ранжирования веб-страниц в поисковых алгоритмах

$$C_i^{PageRank} = \alpha \cdot \sum_j \frac{C_j^{PageRank}}{C_i^{out-deg}} \cdot a_{ij} + (1-\alpha)\frac{1}{n}$$

- Модель PageRank описывает модель поведения случайного пользователя Интернета, который блуждает по сети, переходя со страницы на страницу по гиперссылкам (ребрам), и все такие переходы одинаково вероятны.
- При этом, если страница не имеет ссылок на другие страницы и пользователь попадает на такую страницу, то он не сможет продолжить свое блуждание. Поэтому и добавлено второе слагаемое в формулу центральности: если пользователь попадает на такую страницу (с отсутствием исходящих ссылок), то он случайным образом набирает другой URL-адрес (то есть случайным образом перемещается на другой узел сети, что происходит с одинаковой вероятностью 1/n) и снова продолжает серфинг по Интернету.
- Коэффициент затухания α определяет вероятность продолжения случайного блуждания по сети и его часто берут равным 0,85.

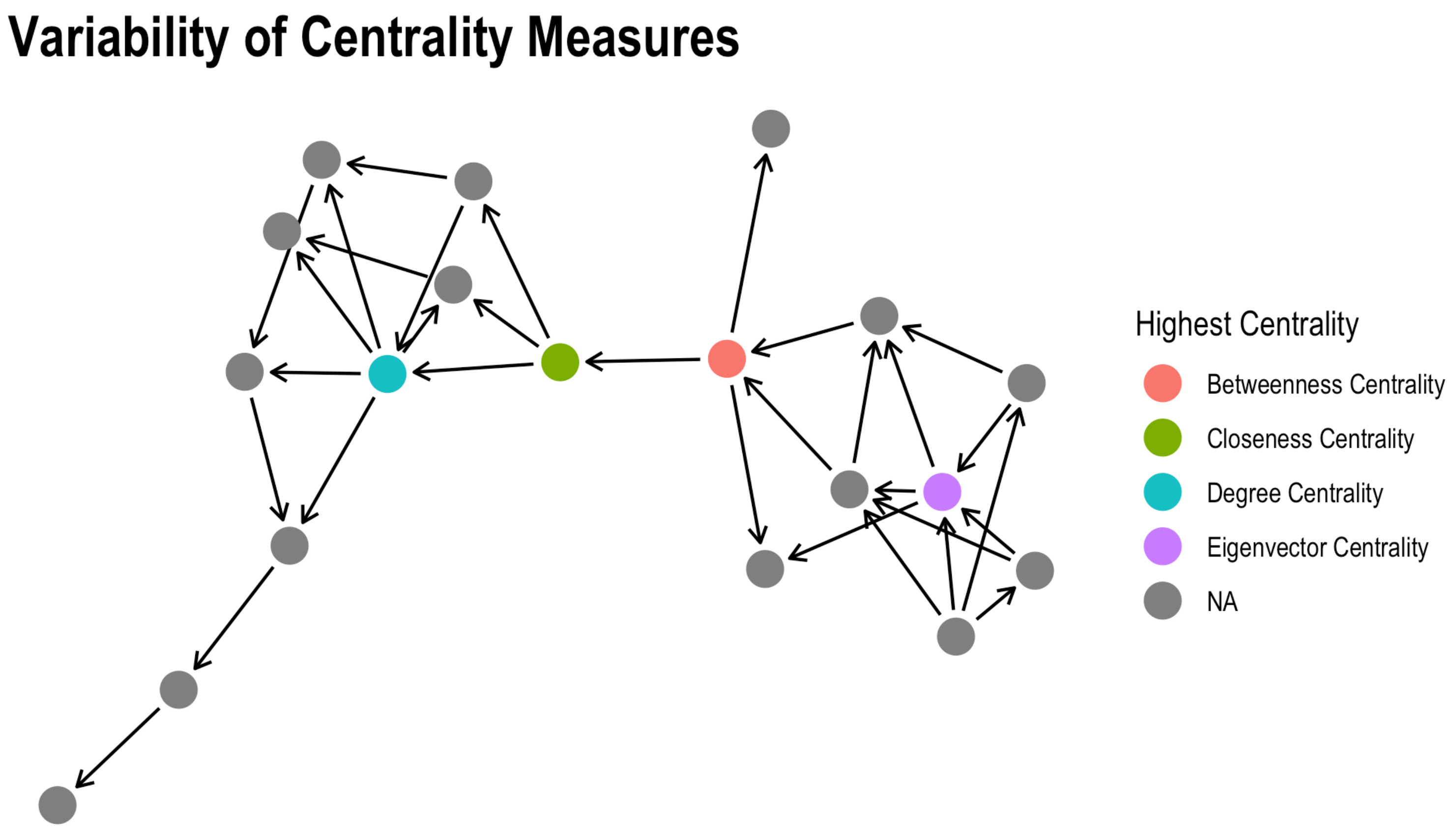
Центральность PageRank (пример)



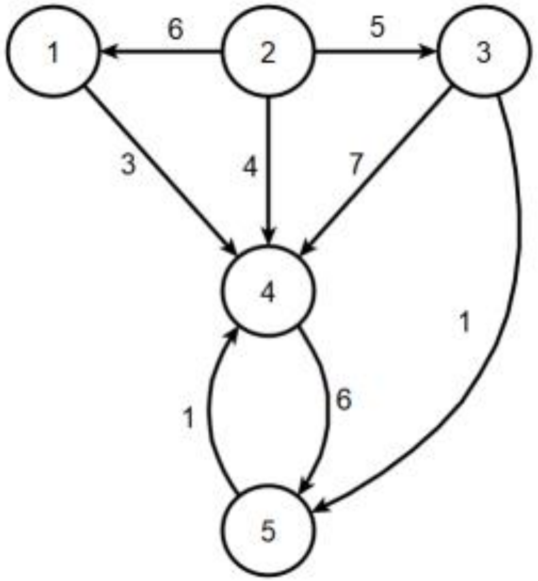
$$
\begin{array}{c c}
 & \begin{array}{ccccccc} A & B & C & D & E & F & G \end{array} \\
\begin{array}{c} A \\ B \\ C \\ D \\ E \\ F \\ G \end{array} &
\begin{pmatrix}
0 & 1 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 & 1 & 0
\end{pmatrix}
\end{array}
$$

i	C_i^{ev}	$C_i^{PageRank}$
A	0,87	0,159
B	0,73	0,17
C	1	0,21
D	0,25	0,07
E	0,50	0,115
F	0,45	0,116
G	0,80	0,160

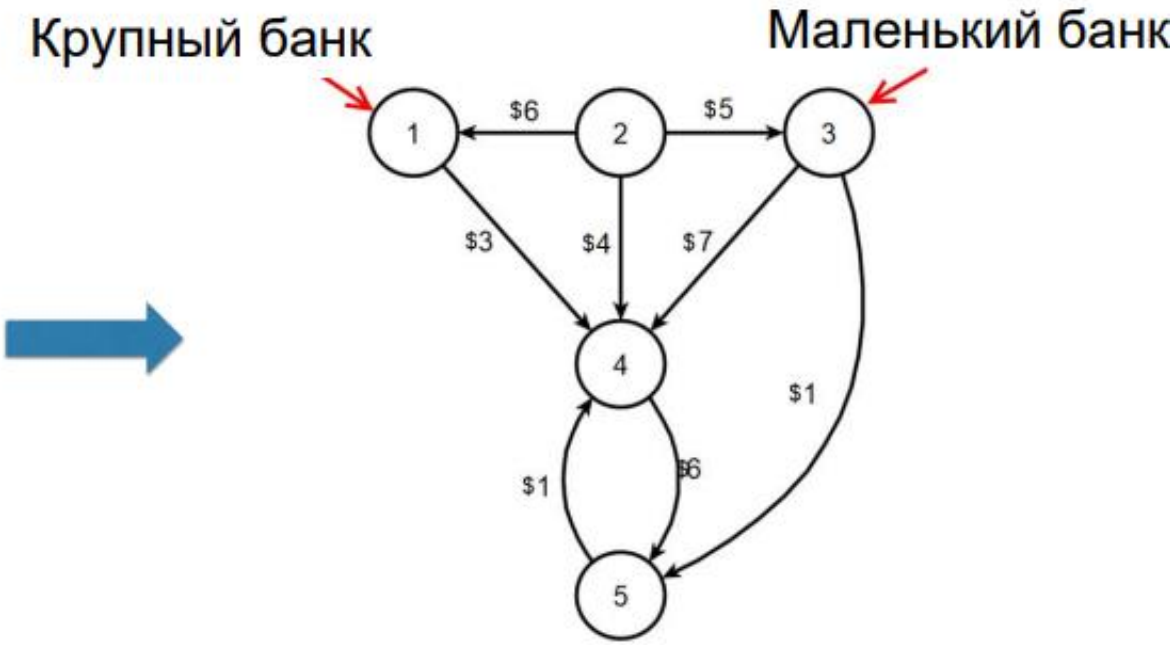
Сравнение мер центральности



ЦЕНТРАЛЬНОСТЬ В СЕТИ: ПРОБЛЕМА



Какая вершина самая важная в сети?



Во многих реальных задачах вершины сети неоднородны
и могут иметь индивидуальные характеристики…

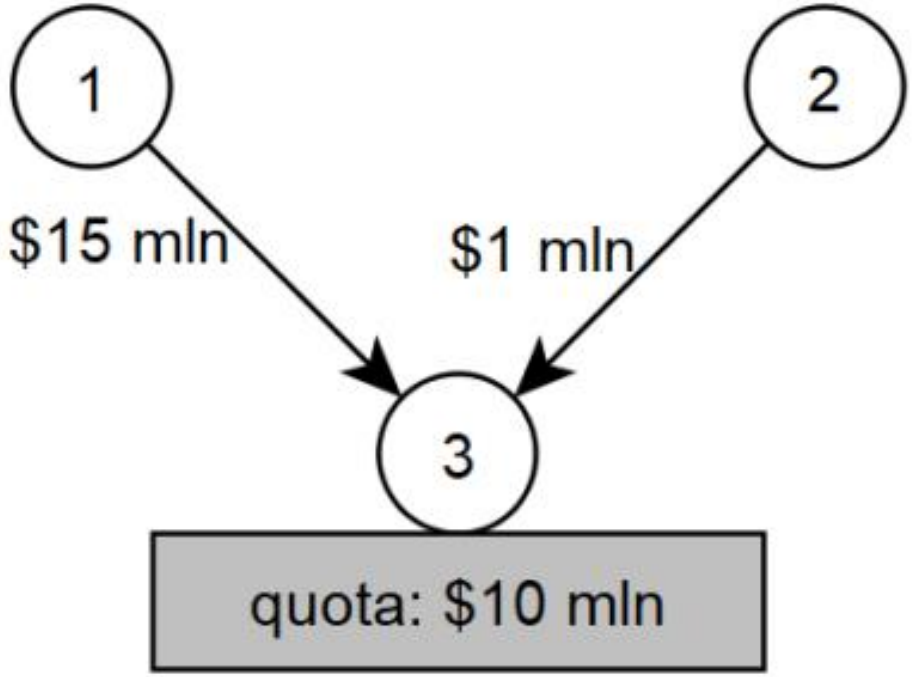
ЦЕНТРАЛЬНОСТЬ В СЕТИ: ПРОБЛЕМА

- Недостатки классических индексов:
- Не учитываются индивидуальные атрибуты вершин;
- Не учитывается групповое влияние вершин;
- Учитываются все связи в сети;
- Не учитываются дальние взаимодействия.

Индекс дальних взаимодействий (LRIC)

- Авторы: Алескеров Ф.Т., Швыдун С.В., Мещерякова Н.Г.

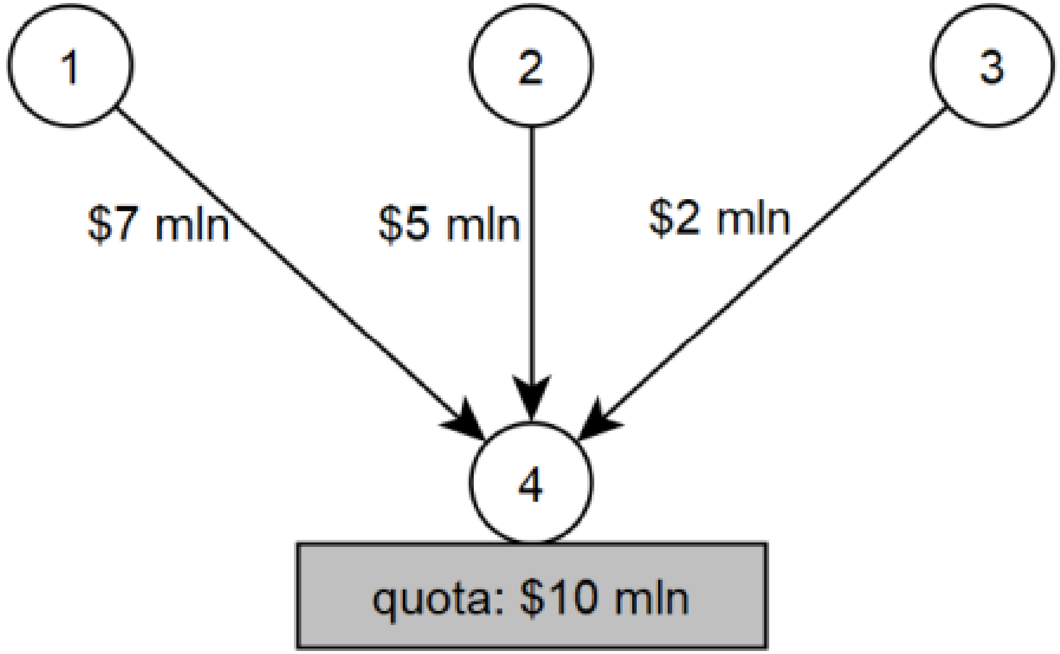
Прямое влияние



Квота – пороговое значение, при превышении которого вершины влияют на другую вершину.

Индекс дальних взаимодействий (LRIC)

Групповое влияние



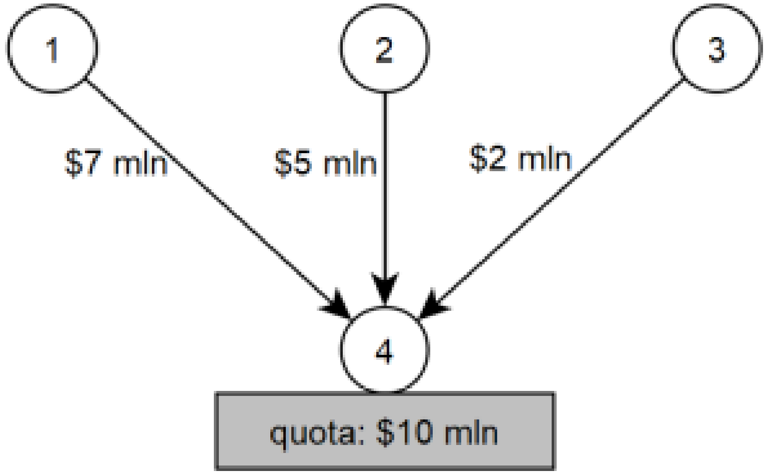
Критическая группа: группа вершин, которая превосходит квоту.

Ключевая игрок: игрок является ключевым, если без него группа становится некритической.

Индекс дальних взаимодействий (LRIC)

Критическая группа: группа вершин, которая превосходит квоту.

Ключевой игрок: игрок является ключевым, если без него группа становится некритической.



Критические группы:

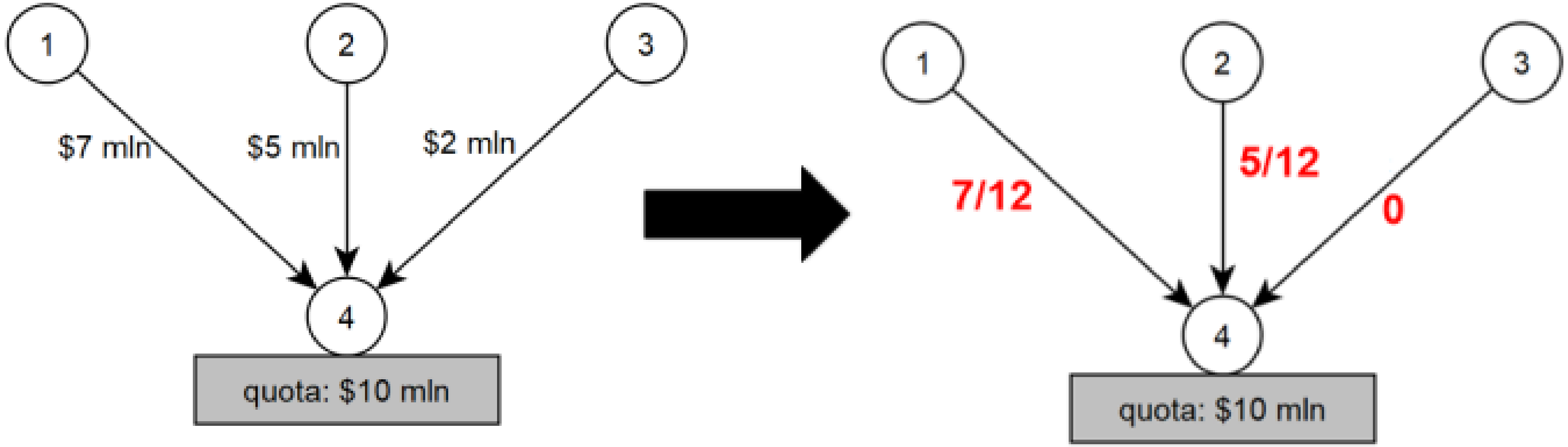
- **{1,2}** ($12 mln >$10 mln)
- **{1,2,3}** ($14 mln >$10 mln);

Ключевые вершины: 1, 2.

Индекс дальних взаимодействий (LRIC)

Критическая группа: группа вершин, которая превосходит квоту.

Ключевой игрок: игрок является ключевым, если без него группа становится некритической.



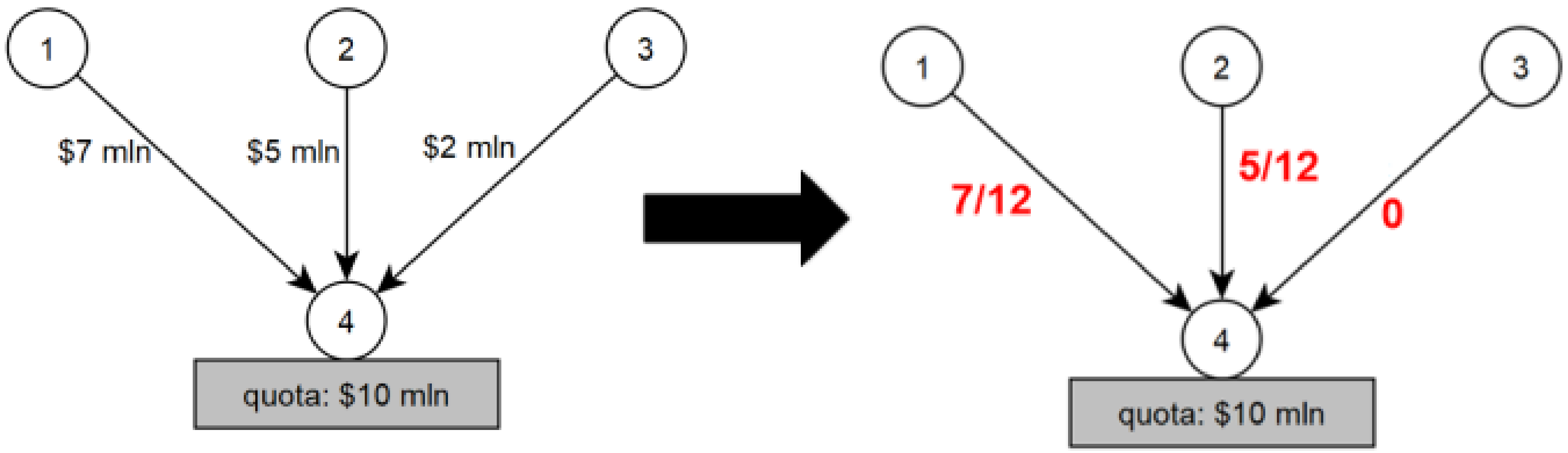
Критические группы:

- **{1,2}** ($12 mln >$10 mln)
- **{1,2,3}** ($14 mln >$10 mln);

Ключевые вершины: 1, 2.

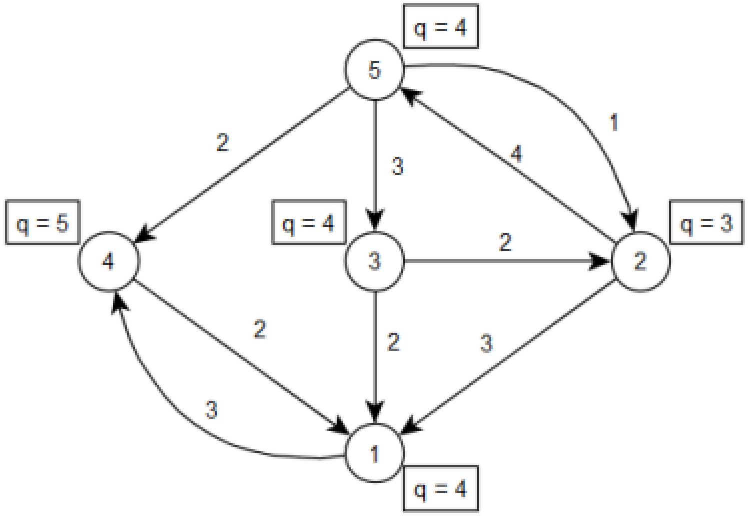
Индекс дальних взаимодействий (LRIC)

- Здесь сначала определяем все критические группы для вершины j, в которых вершина i является ключевой, далее определяем вклад вершины i в суммарный вес этой критической группы (т.е. долю в суммарном весе), и в качестве оценки влияния вершины i на вершину j берем ту оценку вклада, которая является наибольшей. Таким образом, численно считаем влияние в виде наибольшего вклада вершины как ключевого элемента в критической группе среди всех возможных критических групп.
- Если же нет прямой связи от вершины i к вершине j, или вершина i не является ключевой ни в одной из критических групп вершины j, то влияние вершины i на вершину j очевидным образом должно быть равным нулю.



Индекс дальних взаимодействий (LRIC)

- Пример расчета прямого влияния в индексе LRIC

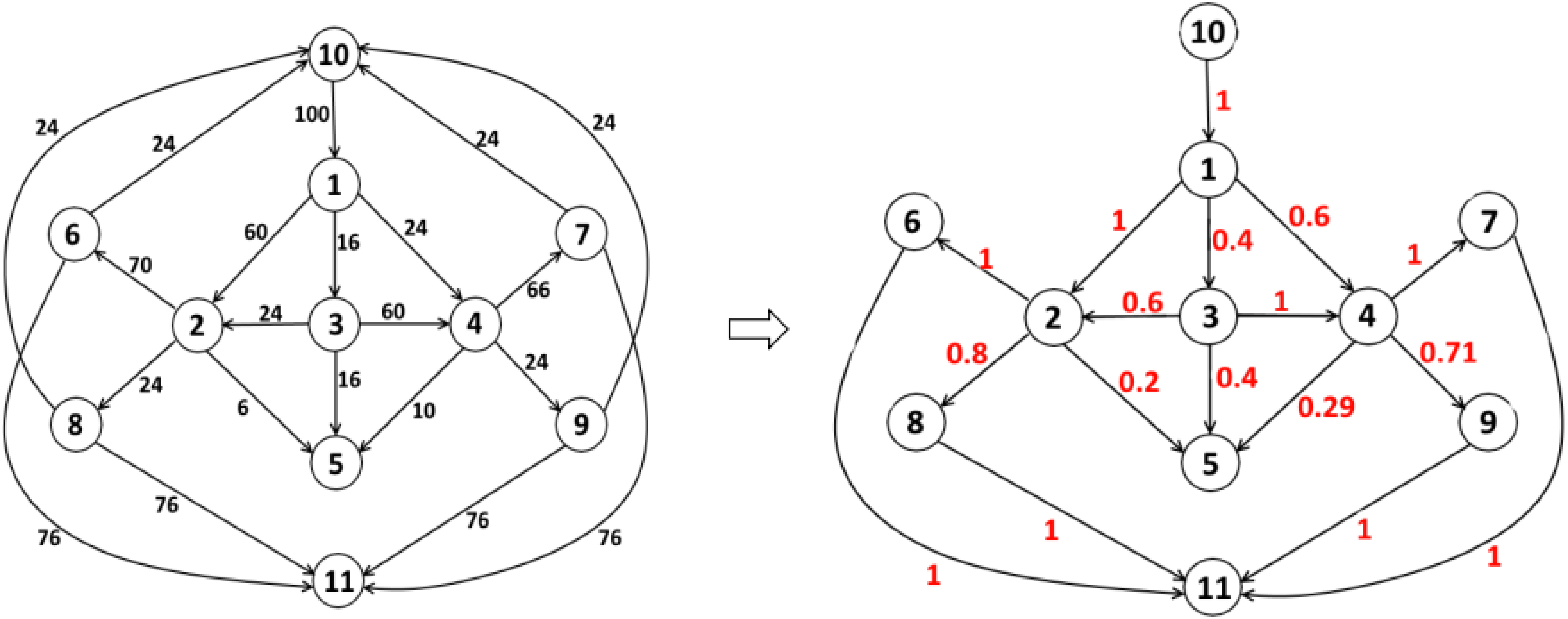


Смежная вершина i	$\Omega_k(j)$, в которой вклад вершины j максимален	Вес группы $\sum_{t\in\Omega_k(j)} w_{ti}$	Прямое влияние c_{ij}^{direct}
		$j=1$	
$\{2\}$	$\{2,3\}$ или $\{2,4\}$	5	$c_{21}^{direct}=3/5$
$\{3\}$	$\{3,4\}$	4	$c_{31}^{direct}=2/4$
$\{4\}$	$\{3,4\}$	4	$c_{41}^{direct}=2/4$
		$j=2$	
$\{3\}$	$\{3,5\}$	3	$c_{32}^{direct}=2/3$
$\{5\}$	$\{3,5\}$	3	$c_{52}^{direct}=1/3$
		$j=3$	
$\{5\}$	$\varnothing$	-	$c_{53}^{direct}=0$
		$j=4$	
$\{1\}$	$\{1,5\}$	5	$c_{14}^{direct}=3/5$
$\{5\}$	$\{1,5\}$	5	$c_{54}^{direct}=2/5$
		$j=5$	
$\{2\}$	$\{2\}$	4	$c_{25}^{direct}=1$

Матрица прямого влияния:

$$\begin{array}{c c}
 & \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} & \begin{pmatrix} 0 & 0 & 0 & 0{,}6 & 0 \\ 0{,}6 & 0 & 0 & 0 & 1 \\ 0{,}5 & 2/3 & 0 & 0 & 0 \\ 0{,}5 & 0 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & 0{,}4 & 0 \end{pmatrix}
\end{array}$$

Индекс дальних взаимодействий (LRIC)



Библиотека в Python: SLRIC

Индекс дальних взаимодействий (LRIC)

Для учета непрямых взаимодействий для оценки влияния вершины i на вершину j по LRIC необходимо выписать все простые пути P_{ij} от i к j и агрегировать влияния каждого этапа этого пути в единую оценку $f(P_{ij})$ интенсивности влияния по непрямому пути. Для этого можно:

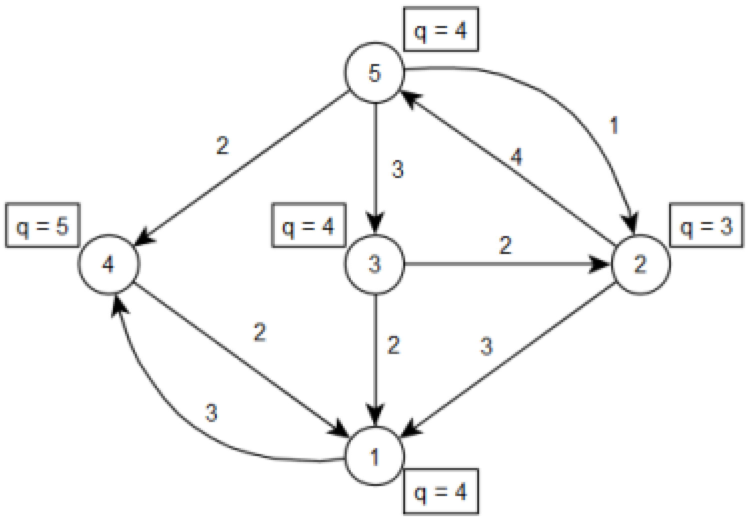
— перемножить оценки прямых влияний $f_{mult}(P_{ij}) = c_{ik_1}^{direct} \times c_{k_1 k_2}^{direct} \times \ldots \times c_{k_{s-1} j}^{direct}$,

— взять минимум $f_{min}(P_{ij}) = min\{c_{ik_1}^{direct}, c_{k_1 k_2}^{direct}, \ldots, c_{k_{s-1} j}^{direct}\}$.

В первом случае можно интерпретировать значения $c_{t_1 t_2}$ как вероятности влияния вершины t_1 на вершину t_2 на этом непрямом пути и $f_{mult}(P_{ij})$ как итоговую вероятность влияния вершины i на вершину j сквозь весь этот путь P_{ij}, а второй вариант формулы основан на идее о том, что влияние вершины i на вершину j сквозь весь путь P_{ij} будет определяться влиянием ребра в самом «узком» месте.

Индекс дальних взаимодействий (LRIC)

- Пример расчета непрямого влияния по пути в индексе LRIC



Есть два непрямых пути от вершины 3 к вершине 1:

— первый $P_{31}^1 : 3 \overset{2/3}{\to} 2 \overset{3/5}{\to} 1$ длины 2,

— второй $P_{31}^2 : 3 \overset{2/3}{\to} 2 \overset{1}{\to} 5 \overset{2/5}{\to} 4 \overset{2/4}{\to} 1$ длины 4.

Выберем первую формулу $f_{mult}(P_{ij})$ агрегирования влияния по одному пути:

$$f_{mult}(P_{31}^1) = f_{mult}(3 \overset{2/3}{\to} 2 \overset{3/5}{\to} 1) = 2/3 \times 3/5 = 2/5,$$

$$f_{mult}(P_{31}^2) = f_{mult}(3 \overset{2/3}{\to} 2 \overset{1}{\to} 5 \overset{2/5}{\to} 4 \overset{2/4}{\to} 1) = 2/3 \times 1 \times 2/5 \times 2/4 = 2/15.$$

Матрица прямого влияния:

$$\begin{array}{c c}
 & \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} & \begin{pmatrix} 0 & 0 & 0 & 0{,}6 & 0 \\ 0{,}6 & 0 & 0 & 0 & 1 \\ 0{,}5 & 2/3 & 0 & 0 & 0 \\ 0{,}5 & 0 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & 0{,}4 & 0 \end{pmatrix}
\end{array}$$

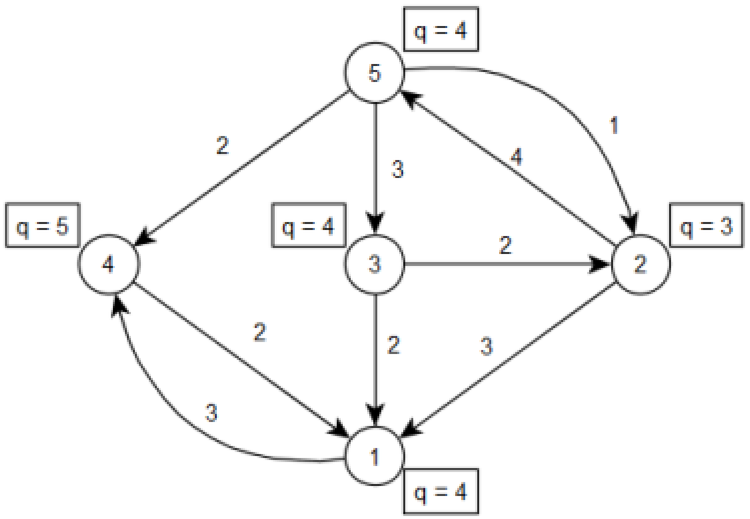
Индекс дальних взаимодействий (LRIC)

Поскольку путей P_{ij} из вершины i в вершину j может быть много (в том числе и прямой путь), то необходимо агрегировать также все оценки влияния этих путей в единое значение индекса полного влияния i на j. Здесь также есть разные варианты агрегирования, зависящие от контекста решаемой задачи. Например, можно аккумулировать интенсивности непрямых влияний как $c_{ij}^{sum} = min(\sum_{P_{ij}} f(P_{ij}, 1))$ (учитывая, что индексы влияния не могут быть более 1), или брать максимум из возможных значений $c_{ij}^{max} = max_{P_{ij}} f(P_{ij})$. Иногда на практике имеет смысл не учитывать слишком длинные пути и ограничиться анализом путей длиной не более 4 (как правило, учет путей большей длины не меняет существенно результаты расчетов).

Если путей из вершины i в вершину j нет, то в оценке полного влияния будем учитывать только прямое влияние вершины i на вершину j.

Индекс дальних взаимодействий (LRIC)

- Пример расчета непрямого влияния между вершинами в индексе LRIC



Есть два непрямых пути от вершины 3 к вершине 1:

— первый $P^1_{31} : 3 \xrightarrow{2/3} 2 \xrightarrow{3/5} 1$ длины 2,

— второй $P^2_{31} : 3 \xrightarrow{2/3} 2 \xrightarrow{1} 5 \xrightarrow{2/5} 4 \xrightarrow{2/4} 1$ длины 4.

Выберем первую формулу $f_{mult}(P_{ij})$ агрегирования влияния по одному пути:

$$f_{mult}(P^1_{31}) = f_{mult}(3 \xrightarrow{2/3} 2 \xrightarrow{3/5} 1) = 2/3 \times 3/5 = 2/5,$$

$$f_{mult}(P^2_{31}) = f_{mult}(3 \xrightarrow{2/3} 2 \xrightarrow{1} 5 \xrightarrow{2/5} 4 \xrightarrow{2/4} 1) = 2/3 \times 1 \times 2/5 \times 2/4 = 2/15.$$

Для расчета полного влияния вершины 3 на вершину 1 надо учесть прямое влияние $c^{direct}_{31} = 2/4$ (рассчитанное в предыдущем примере) и две оценки интенсивности влияния по двум непрямым путям $f_{mult}(P^1_{31}) = 2/5$ и $f_{mult}(P^2_{31}) = 2/15$. Выберем вторую формулу $c^{max}_{31} = max_{P_{ij}} f(P_{ij}) = max\{2/4, 2/5, 2/15\} = 1/2$.

Оставляем читателю самому рассчитать непрямые и полные влияния остальных вершин этой сети по тем же формулам, что и для вершин 3 и 1. Матрица полного влияния по модели LRIC выглядит следующим образом:

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0{,}6 & 0 \\ 0{,}6 & 0 & 0 & 0{,}6 & 1 \\ 0{,}5 & 2/3 & 0 & 0{,}3 & 2/3 \\ 0{,}5 & 0 & 0 & 0 & 0 \\ 0{,}2 & 1/3 & 0 & 0{,}4 & 0 \end{pmatrix} \end{array}$$

Матрица прямого влияния:

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0{,}6 & 0 \\ 0{,}6 & 0 & 0 & 0 & 1 \\ 0{,}5 & 2/3 & 0 & 0 & 0 \\ 0{,}5 & 0 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & 0{,}4 & 0 \end{pmatrix} \end{array}$$

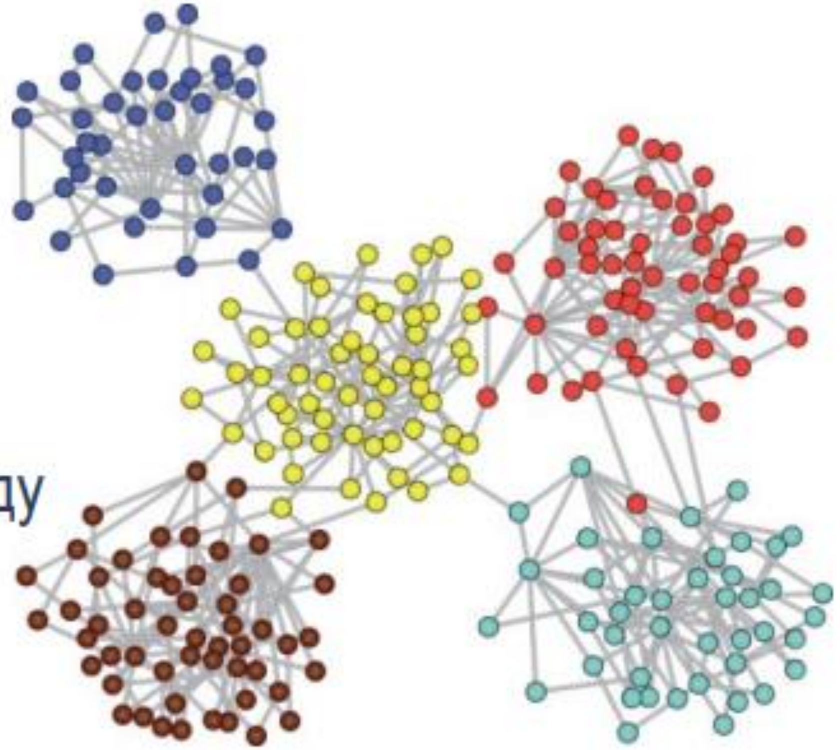
Модели выявления сообществ

Цель: *найти группы вершин, которые в некотором смысле более похожи друг на друга, чем на другие вершины сети.*

Модели выявления сообществ

Основные особенности:

- *Связность: Узлы внутри одного сообщества связаны имеют связи друг с другом.*
- *Компактность*: члены сообщества находятся на небольшом расстоянии друг от друга;
- *Плотность ребер*: высокая частота ребер внутри сообщества;
- *Отделимость*: более высокая частота ребер между членами сообщества по сравнению с не членами.

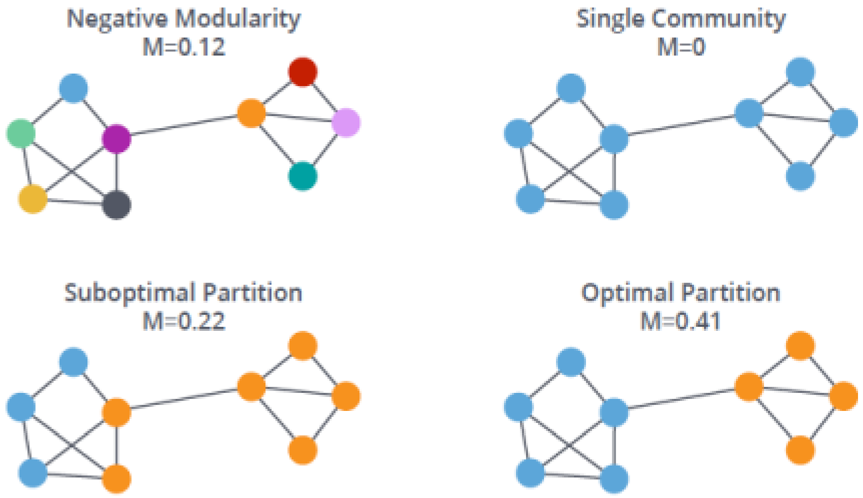


Модели выявления сообществ

- Методы:
  - Модулярность (Modularity). Идея: сравнить долю ребер внутри сообщества по сравнению с ожидаемой долей ребер в случайном графе, имеющем идентичное степенное распределение.

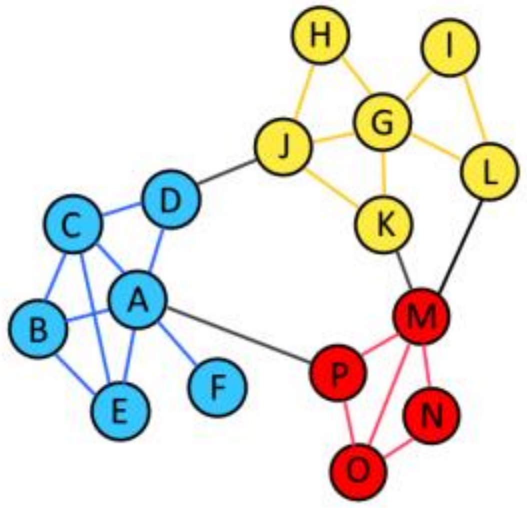
$$M = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j) \to max$$

- $\delta(c_i, c_j) = 1$, если вершины *i* и *j* находятся в одном сообществе и $\delta(c_i, c_j) = 0$ в ином случае;
- k_i – степень вершины *i*.
- *m – общее число ребер в сети.*



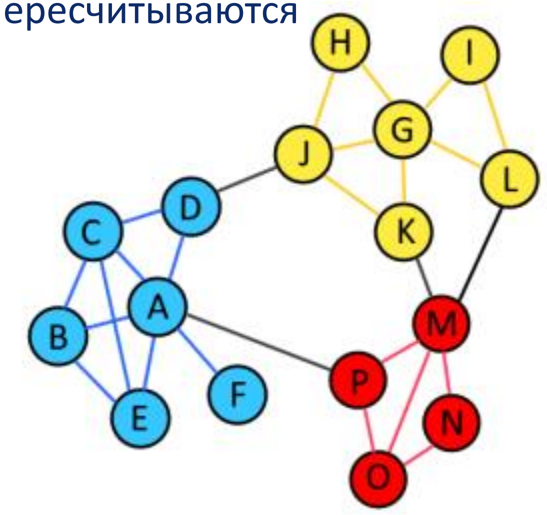
Модели выявления сообществ

- Методы:
  - Модулярность (Modularity). Идея: сравнить долю ребер внутри сообщества по сравнению с ожидаемой долей ребер в случайном графе, имеющем идентичное степенное распределение.
  - Алгоритм «label.propagation.community». Идея: первоначально каждая вершина сети имеет свой собственный ярлык, затем начинается итеративный процесс и на каждой итерации часть вершин принимает ярлыки своих «соседей», а именно тот ярлык, который более представлен среди соседей данной вершины. Процесс продолжается до тех пор, пока итерации не прекратятся



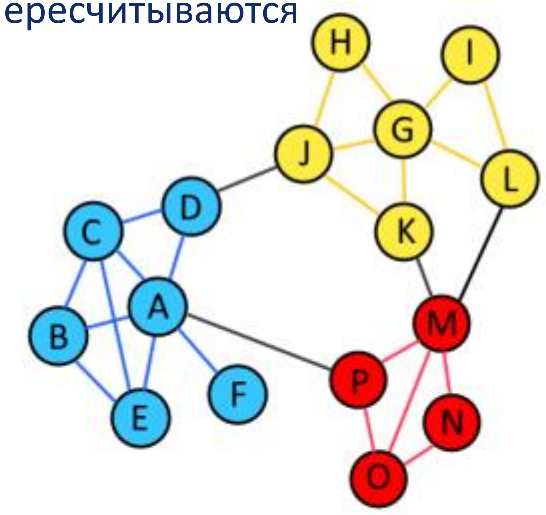
Модели выявления сообществ

- Методы:
  - Модулярность (Modularity). Идея: сравнить долю ребер внутри сообщества по сравнению с ожидаемой долей ребер в случайном графе, имеющем идентичное степенное распределение.
  - Алгоритм «label.propagation.community». Идея: первоначально каждая вершина сети имеет свой собственный ярлык, затем начинается итеративный процесс и на каждой итерации часть вершин принимает ярлыки своих «соседей», а именно тот ярлык, который более представлен среди соседей данной вершины. Процесс продолжается до тех пор, пока итерации не прекратятся
  - Промежуточность ребер (Edge betwenneess). Идея: все пути между вершинами различных сообществ проходят через крайне небольшое число ребер, которые соединяют данные сообщества. Промежуточность ребра– число кратчайших путей, проходящих через данное ребро. Происходит итеративное исключение ребер с наибольшей промежуточностью, и после каждого исключенного ребра значения промежуточности пересчитываются



Модели выявления сообществ

- Методы:
  - Модулярность (Modularity). Идея: сравнить долю ребер внутри сообщества по сравнению с ожидаемой долей ребер в случайном графе, имеющем идентичное степенное распределение.
  - Алгоритм «label.propagation.community». Идея: первоначально каждая вершина сети имеет свой собственный ярлык, затем начинается итеративный процесс и на каждой итерации часть вершин принимает ярлыки своих «соседей», а именно тот ярлык, который более представлен среди соседей данной вершины. Процесс продолжается до тех пор, пока итерации не прекратятся
  - Промежуточность ребер (Edge betwenneess). Идея: все пути между вершинами различных сообществ проходят через крайне небольшое число ребер, которые соединяют данные сообщества. Промежуточность ребра– число кратчайших путей, проходящих через данное ребро. Происходит итеративное исключение ребер с наибольшей промежуточностью, и после каждого исключенного ребра значения промежуточности пересчитываются
  - Спектральная кластеризация (Spectral clustering);
  - Кластеризация на основе блужданий в графах (Walktrap methods);



Модели сетевого анализа

- Типы взаимоотношений:
  - Внешние заимствования;
  - Международная миграция;
  - Экспорт / импорт продуктов питания;
  - Ранжирование журналов;
  - Обмен студентами;
  - Международные конфликты;
  - Экспорт технологий;
  - Банковские долги;
  - Многие другие.

ВЫСШАЯ • ШКОЛА • ЭКОНОМИКИ