УДК 316.023.6

СКРЫТЫЕ СВЯЗИ В СЛОЖНЫХ СЕТЯХ: МЕТОД СОПРОТИВЛЕНИЙ

А. А. Снарский[1,2], Д.И. Зоринец[1], Д. В. Ландэ[1,2]
[1]Национальный технический университет Украины «Киевский политехнический институт», проспект Победы, 37, 03056 Киев
[2]Институт проблем регистрации информации НАН Украины ул. Н. Шпака, 2, 03113 Киев

Рассмотрена характеристика, описывающая связь узлов в сложной сети, показывающая насколько тесно они связаны между собой. Предложен и обоснован алгоритм расчета такой характеристики, определяемой как «скрытая связь». Приведены примеры расчёта скрытых связей для нескольких социальных сетей.

Введение

Среди множества характеристик сложных сетей (Complex Networks) [1-3] большая часть характеризует узлы и существующие между ними связи, однако существует класс задач, которые изучают отсутствие связей, такие как, например, предсказание появления связи между узлами (Link Prediction)[4] и нахождения скрытых связей в сети (Hidden Links)[5].

В данной работе рассматривается характеристика, описывающая взаимодействие двух узлов, не связанных напрямую друг с другом.

Качественное содержание этой характеристики можно пояснить на примере социальной сети – группе людей, некоторые из которых связаны друг с другом. В такой сети возможна ситуация, когда два человека (узла сети), не связанные непосредственно друг с другом, имеют много общих друзей. Другими словами, связи в один шаг (прямой связи) между ними нет, но присутствует несколько связей длинной больше двух (через одного, двух и более людей). Можно предположить, что между такими двумя людьми на самом деле существует прямая связь, не выявленная по некоторым причинам при исследовании сети. Возможно, эта связь искусственно скрывается. Чем больше количество непрямых связей между данной парой узлов, тем вероятней присутствие между ними такой «виртуальной» скрытой связи.

Цель работы

Цель данной работы – предложить, обосновать и привести примеры расчёта количественной оценки характеристики «скрытая связь» (*HL*).

В работе будет дан метод расчета и приведен пример для простого графа, после чего будет представлена эталонная социальная сеть, так называемый «клуб карате»[6]

Нахождение скрытых связей

Для нахождения скрытых связей в сети представим ее в виде электрической цепи, каждая связь в которой обладает единичным сопротивлением. Тогда между любыми двумя узлами i и j возможно найти сопротивление. Такая характеристика известна как резистивное расстояние (Resistance Distance) [3].

Положим, что сопротивление каждой связи равно:

$$R = \frac{1}{a_{ij}}, \tag{1}$$

где a_{ij} - элемент матрицы смежности сети. Тогда резистивное расстояние между двумя узлами i и j можно получить, воспользовавшись матрицей Лапласа:

$$R_{ij} = \frac{L_{(i)}}{L_{(i,j)}}, \tag{2}$$

где $L_{(i)}$ - алгебраическое дополнение матрицы Лапласа, а $L_{(i,j)}$ - алгебраическое дополнение второго порядка, то есть определитель матрицы, полученный из матрицы Лапласа путем вычеркивания двух строк и столбцов i и j. Таким образом из матрицы резистивных расстояний можно получить матрицу кондактансов для каждой пары узлов i и j:

$$G_{ij} = \frac{1}{R_{ij}}, \tag{3}$$

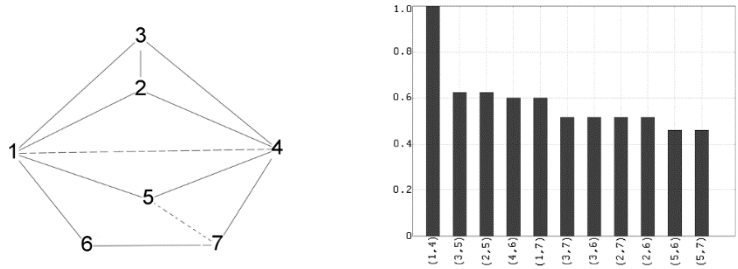
Если для некоторой пары узлов i и j, напрямую не связанных между собой, G_{ij} не меньше некоторого значения G_c:

$$G_{ij} \geq G_c, \tag{4}$$

то предполагается, что между узлами i и j присутствует HL.

Рассмотрим простой пример – расчет G_{ij} для графа с семью узлами изображенного на рис. 1.

Для данного графа скрытая связь между узлами 1 и 4 является самой очевидной, а связь $(5,7)$ самой «слабой», что явно видно из рис. 1а. Очевидно, что скрытая связь $(1,4)$ обладает самым большим значением параметра G_{ij} в силу наибольшего количества параллельных связей между узлами 1 и 4.

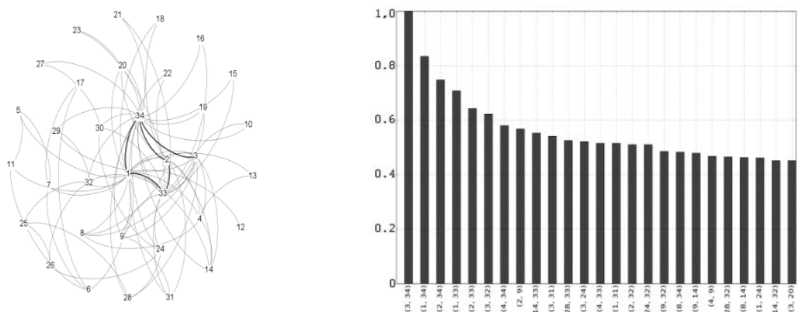


а) б)

Рис 1. Граф и ранжированное распределение

а) граф, состоящий из семи узлов, сплошными линиями показаны связи, пунктирными –скрытые связи, с наибольшим значением $G_{1,4}$ и наименьшим $G_{5,7}$; б) значения G_{ij} для всех пар узлов, не связанных напрямую друг с другом: ранжированное распределение G_{ij}, нормированное на наибольшее значение (т.о. $G_{1,4}=1$). Внизу указаны пары узлов для G_{ij}.

Рассмотрим реальную сеть «клуб каратэ» [6] – социальную сеть дружеских отношений между 34 членами клуба каратэ в университете США в 1970-е годы.



а) б)

Рис 2. Сеть «клуб карате» и зависимость параметра G_{ij}

а) геометрическое представление сети «клуб каратэ» (связанные узлы) и пять HL - жирные линии, с наибольшим значением параметра G ; б) значения G_{ij} для всех пар узлов, не связанных напрямую друг с другом: ранжированное распределение G_{ij}, нормированное на наибольшее значение (т.о. $G_{3,34}=1$). Внизу указаны пары узлов для G_{ij}.

Выводы

Данный метод позволяет идентифицировать неявные или скрытые связи в сложной сети. Таким образом, можно обнаружить целые подгруппы, связанные скрытыми связями и образующие подсеть. К понятию HL близко примыкает понятие регулярной эквивалентности [1] и проблема *Link Prediction* [4]. Регулярная эквивалентность двух узлов определяется схожестью соседей у этих узлов. *Link Prediction* характеризует вероятность возникновения связи между данными узлами при росте сети.

Существует несколько методов оценки вероятности возникновения связи между двумя узлами i и j, которые не связаны друг с другом [4], например, определяется количество общих соседних узлов (структурная эквивалентность) $score(i, j)=\left|\Gamma(i)\cap\Gamma(j)\right|$, где $score(i, j)$ — это величина, по которой происходит оценка вероятности возникновения связи, а $\Gamma(v)$ — множество узлов, длина пути от которых к узлу v равна 1. И чем величина $score(i, j)$ больше, тем выше вероятность появления связи между узлами i и j.

Еще один метод состоит в вычислении коэффициента Жаккара $score(i, j)=\left|\Gamma(i)\cap\Gamma(j)\right|/\left|\Gamma(i)\cup\Gamma(j)\right|$, где $score(i, j)\in\left[0,1\right]$.

Следует отметить, что данные оценки учитывают только ближайших соседей, т.е. узлы, длина пути между которыми больше или равна двум, будут иметь $score(i, j)=0$. Еще два метода оценки вероятности возникновения связи — произведение степеней узлов $score(i, j)=\left|\Gamma(i)\right|\left|\Gamma(j)\right|$ и величина, обратная самой короткой длине пути между узлами i и j $score(i, j)=1/P(i, j)$.

Для предсказания появления связей в иерархических случайных сетях существуют также алгоритмы, основанные на случайном блуждании по такой сети, реализованные методом Монте-Карло[4]. Из результирующей выборки всех возможных связей, ранжированных по вероятности возникновения, только верхний 1% можно считать достаточно точно предсказанным.

В [7] рассмотрен метод определения так называемой, сети доверия преступников. Алгоритм нахождения такой сети основан на вычислении кратчайших путей между известными узлами–преступниками, самыми влиятельными узлами в сети и некоторыми средними узлами. В предложенном алгоритме необходима начальная выборка из узлов-преступников, и учитываются только кратчайшие пути между узлами.

Литература

1. S. N. Dorogovtsev; J. F. F. Mendes, Evolution of networks // Advances in Physics, 51, 2002.
2. M. E. J. Newman. Networks: an Introduction // Oxford University Press, Computers, 2010.
3. E.Estrada. The Sructure of Conplex Netwirks // Oxford University Press, 2011.
4. Aaron Clauset, Cristopher Moore, M.E.J. Newman. Hierarchical structure and the prediction of missing links in networks // arXiv:0811.0484 [stat.ML], 2008.
5. M.I. Zhenirovskyy, D.V. Lande, A.A. Snarskii, Detection Implicit Links and G-betweenness. // arXiv:1008.4073 [cond-mat.dis-nn], 2010.
6. W. W. Zachary, An information flow model for conflict and fission in small groups // Journal of Anthropological Research 33, pp.452-473, 1977.
7. Pritheega Magalingam, Asha Rao, Stephen Davis. Identifying a Criminal's Network of Trust // arXiv:1503.04896 [cs.SI], 2015.