Линейна алгебра и анализ на социални мрежи

„Приятелският парадокс” и неравенството на Коши-Шварц

- „Приятелският парадокс” е формулиран за пръв път през 1991 г. от американския социолог Скот Фелд в публикацията му *Why Your Friends Have More Friends Than You Do* и гласи, че средно хората имат по-малко приятели, отколкото техните приятели.

- През 2011 г. е проверена валидността му за Facebook - установено е, че средно един потребител в мрежата има 245 приятели, докато средният брой приятели на приятелите на всеки потребител е 359.

- Валидността му е установена и за над 98% от потребителите на Twitter - потребителите, които даден потребител следва, имат средно повече последователи от самия него и т.н.

- Парадоксът е обобщен H. Eom, H. Jo и D. Higham за някои характеристики на потребители в социална мрежа, като например престиж (значимост) и други характеристики, за които съществува положителна корелация с броя приятели.

На какво се дължи парадоксът?

- Когато даден потребител има много приятели, той допринася до броя на приятелите на приятелите на много потребители в мрежата.
- Освен това е по-вероятно да си приятел на потребител с много на брой приятели, отколкото на потребител с малко приятели, т.е. в приятелския ти кръг е по-вероятно да има потребители със средно повече приятели от твоите.

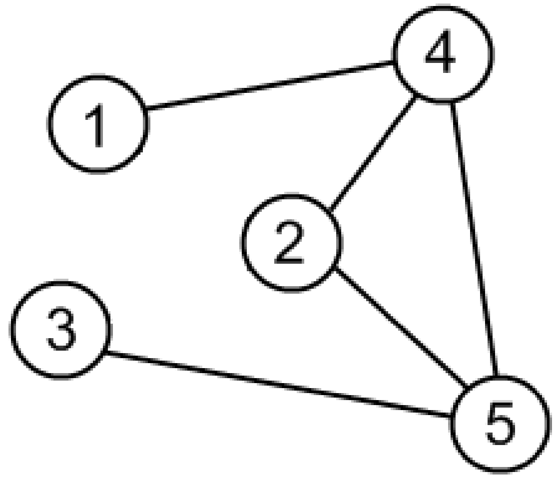
Математически модел на социална мрежа

Често за математически модел на социална мрежа се използва структурата граф.

Граф е математическа структура, чрез която се описват взаимоотношения между обекти. Всеки граф G се състои от две множества - множество на върховете V (обектите) и множество на ребрата E (връзките между обектите).

Неориентиран граф - ребрата нямат посока, изобразяват се като отсечки.

Ориентиран граф - ребрата имат посока, изобразяват се като стрелки (насочени отсечки).



№ на връх	брой приятели	брой приятели на приятелите
1	1	3
2	2	3, 3
3	1	3
4	3	1, 2, 3
5	3	1, 2, 3
средно	$10/5 = \mathbf{2}$	$24/10 = \mathbf{2{,}4}$

Степен на връх се нарича броят ребра, с които е свързан даден връх. Това е и броят на върховете, които се наричат *съседни* на дадения връх, в случай, че в графа няма паралелни ребра (повече от едно ребро, свързващо два върха; мултиграф) и примки (ребра с краища един и същи връх).

Нека с d_i означим степента на върха с пореден номер i. Тъй като приятелството между два върха означаваме с едно ребро между тях, то d_i е броят на приятелите на върха i.

В терминологията на теорията на графите приятелският парадокс означава, че средната степен на върховете не надвишава средната степен на съседните им върхове.

Ще разгледаме вектора d с координати d_i, т.е. вектора, съдържащ броя на приятелите на всеки връх $d = (d_1, d_2, ..., d_n)$.

Освен евклидовата норма на произволен вектор $v = (x_1, x_2, ..., x_n)$, т.е.

$$||v|| = \sqrt{v^2} = \sqrt{x_1^2 + x_2^2 + ... + x_n^2},$$

ще разглеждаме и още една векторна норма, известна като *единична норма*, която се определя чрез равенството

$$||v||_1 = |x_1| + |x_2| + ... + |x_n|.$$

Ако за всички координати на v е изпълнено $x_i \geq 0$, то

$$||v||_1 = x_1 + x_2 + ... + x_n.$$

В тези означения, средният брой приятели на върховете в граф се пресмята чрез формулата

$$\text{среден брой приятели} = \frac{d_1 + d_2 + ... + d_n}{n},$$

където n е общият брой върхове на графа. Горното равенство може да се запише още така

$$\text{среден брой приятели} = \frac{||d||_1}{n}.$$

Сега нека изведем формула за пресмятането на средния брой приятели на приятелите на върховете в граф. Всеки връх i участва в приятелския кръг на d_i на брой върха и при всяко такова участие допринася с d_i на брой приятели. Тогава сумата на броя на приятелите на приятелите на върховете в графа е равен на скаларния квадрат $d.d = d^2 = ||d||^2$.

Следователно средният брой приятели на приятелите на връх от графа е

$$\text{среден брой приятели на приятели} = \frac{d^2}{||d||_1}.$$

За разглеждания от нас пример

$$d = (1, 2, 1, 3, 3), \qquad n = 5,$$
$$||d||_1 = 1 + 2 + 1 + 3 + 3 = 10,$$
$$d^2 = ||d||^2 = 1^2 + 2^2 + 1^2 + 3^2 + 3^2 = 24.$$

Тогава

$$\text{среден брой приятели} = \frac{||d||_1}{n} = \frac{10}{5} = 2.$$

$$\text{среден брой приятели на приятели} = \frac{d^2}{||d||_1} = \frac{24}{10} = 2,4.$$

Парадоксът на приятелите в математически означения е

$$\frac{||d||_1}{n} \leq \frac{d^2}{||d||_1}.$$

Ще покажем, че това неравенство е следствие от неравенството на Коши-Шварц.

Теорема *(Неравенство на Коши–Буняковски–Шварц) За всеки два вектора* $\mathbf{u}$ *и* $\mathbf{v}$ *в реално евклидово векторно пространство е в сила неравенството*

$$(\mathbf{u}\mathbf{v})^2 \leq \mathbf{u}^2\mathbf{v}^2.$$

Равенството се достига, точно когато $\mathbf{u}$ *и* $\mathbf{v}$ *са линейно зависими.*

След кореннуване, неравенството е еквивалентно на

$$-||\mathbf{u}||.||\mathbf{v}|| \leq \mathbf{u}\mathbf{v} \leq ||\mathbf{u}||.||\mathbf{v}||$$

Ние ще използваме само тази част на неравенството

$$|uv| \;\leq\; ||u||.||v||,$$

което за вектори u и v с координати положителни числа е еквивалентно на

$$uv \;\leq\; ||u||.||v||$$

Нека u е n-мерен вектор, чиито координати са само единици, т.е. $u = (1, 1, ..., 1)$ и $v = (v_1, v_2, ..., v_n)$, $v_i \geq 0$. Тогава $||u|| = \sqrt{n}$ и $uv = v_1 + v_2 + ... + v_n = ||v||_1$. След заместване в горното неравенство получаваме

$$||v||_1 \;\leq\; \sqrt{n}||v|| \quad \Leftrightarrow \quad ||v||_1^2 \;\leq\; nv^2 \quad \Leftrightarrow \quad \frac{||v||_1}{n} \;\leq\; \frac{v^2}{||v||_1},$$

в случай, че $||v||_1 \neq 0$. Ако в последното получено неравенство заместим v с d, получаваме приятелския парадокс

$$\frac{||d||_1}{n} \;\leq\; \frac{d^2}{||d||_1}.$$

Равенството в горното неравенство се достига, точно когато се достига в неравенството на Коши-Шварц, а именно в случая, когато u и v са колинеарни вектори. За приятелския парадокс това е случаят, когато $u = (1, 1, ..., 1)$ и d са колинеарни, т.е. всички върхове имат равни степени (*регулярен граф*).

Доказателство на неравенството на Коши-Шварц за реални вектори.

Нека u и v са вектори от реалното векторно пространство V. Случаят $u = o$ е тривиален, тъй като тогава $uv = ov = 0$ и $||u||.||v|| = 0.||v|| = 0$ (понеже $||o|| = 0$), т.е. е неравенството е изпълнено $(0 \leq 0)$.

Нека $u \neq o$ и да разгледаме вектора $cu - v$, където $c \in \mathbb{R}$. По-точно нека разгледаме скаларния квадрат на този вектор, за който знаем, че задължително е неотрицателно число, т.е. $(cu - v)^2 \geq 0$. Съгласно свойствата на скаларното произведение този скаларен квадрат е равен на

$$(cu - v)(cu - v) = c^2u^2 - 2c(uv) + v^2 = ||u||^2c^2 - 2(uv)c + ||v||^2.$$

Следователно изразът $||u||^2c^2 - 2(uv)c + ||v||^2 \geq 0$. Нека забележим, че този израз е квадратен тричлен относно реалната променлива c. Означаваме $f(c) = ||u||^2c^2 - 2(uv)c + ||v||^2$.

Освен това квадратният тричлен $f(c)$ приема само неотрицателни стойности за всяка реална стойност на променливата си c, т.е. $f(c) \geq 0$. Последното условие е еквивалентно на условието $D \leq 0$ за дискриминантата на тричлена, т.е.

$$D = (uv)^2 - ||u||^2.||v||^2 \leq 0,$$

което е еквивалентно на неравенството на Коши-Шварц

$$(uv)^2 \leq ||u||^2.||v||^2$$

Да видим кога се достига равенството в горното нестрого неравенство. Равенството е еквивалентно на $D = 0$, което от своя страна означава, че разглежданият квадратен тричлен $f(c)$ има двоен реален корен

$$c_0 = \frac{2(uv)}{2||u||^2} = \frac{uv}{||u||^2}.$$

За тази стойност на $c = c_0$ имаме $f(c_0) = 0$. Но да си припомним, че $f(c) = (cu - v)^2$, следователно $f(c_0) = (c_0 u - v)^2 = 0$, което е възможно единствено в случай, че $c_0 u - v = o$. Последното равенство е еквивалентно на $v = c_0 u$, където $c_0 = \frac{uv}{||u||^2}$, т.е. векторите u и v са колинеарни (с коефициент на колинеарност c_0).

Литература

1. H. Anton, C. Rorres, *Elementary Linear Angebra (applications version)*, 11th ed., Wiley, 2014, ISBN 978-1-118-43441-3.

2. C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM, 2000.

3. S. L. Feld, *Why Your Friends Have More Friends Than You Do*, The American Journal of Sociology, Vol. 96, No. 6 (May, 1991), 1464–1477.

4. D. J. Higham, *Centrality-friendship paradoxes: when our friends are more important than us*, Journal of Complex Networks (2019) 7, 515–528.