Attribute Driven Active Community Search

Badhan Chandra Das, Md Shoaib Ahmed and Md Musfique Anwar
Jahangirnagar University
Email:badhan0951@gmail.com, shoaibmehrab011@gmail.com, manwar@juniv.edu

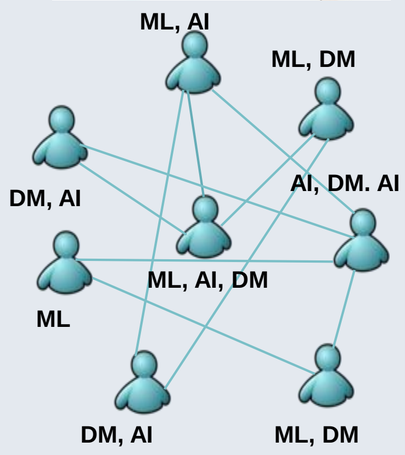
Background

Community

- A **group** of users who are **closely connected** to each other.
- **Interact** with each other more than with those outside the group.

Attributed Graph

- **Rich** attributed nodes with properties.



Classified

i. Community Detection.
ii. Community Search.

Our Observations

- Find attribute-oriented user **active** community.
- **Quantify** users' activeness with regards to attributes.

Motivation and Problem Formulation

Motivation

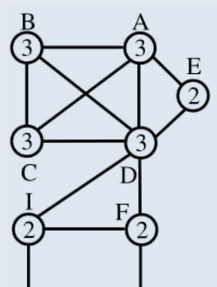
Previous approaches

i. paid **less attention** to the **user interests** and **activities** on different attributes.
ii. **Failed** to search active communities.

Our approaches

i. **Different degree** of activeness for different attributes.
ii. Find an active community according **user interests**.

Problem Formulation



(a) Social Graph

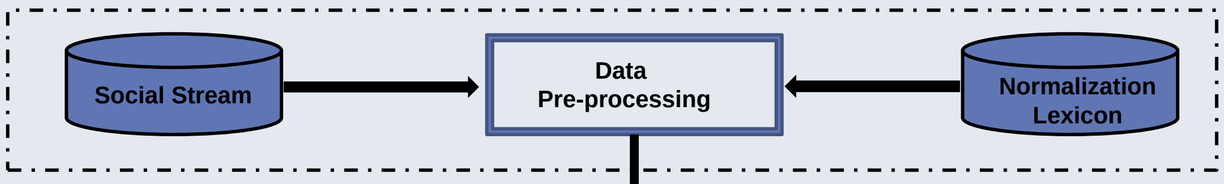
	a_0	a_1
A	5	2
B	2	0
C	3	0
D	4	4
E	5	1
F	0	7
G	2	1
H	1	2
I	0	7

(b) Action log

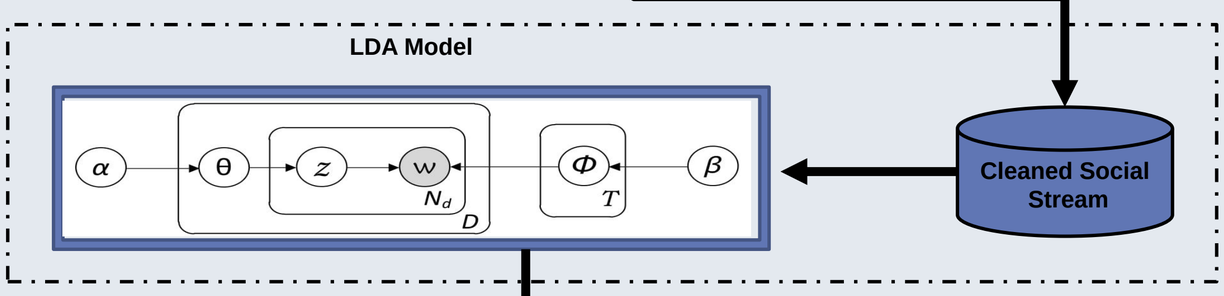
Query	Community ($\mathcal{C}_q$)
$Q=\{D,a_0\}, k=2, \gamma=1$	{A, B, C, D, E}
$Q=\{D,a_0\}, k=3, \gamma=1$	{A, B, C, D}
$Q=\{D,a_1\}, k=2, \gamma=1$	{A, D, E, F, G, H, I}
$Q=\{D,a_1\}, k=3, \gamma=1$	{ }
$Q=\{D,a_0,a_1\}, k=2, \gamma=1$	{A, B, C, D, E, F, G, H, I}
$Q=\{D,a_0,a_1\}, k=3, \gamma=1$	{A, B, C, D}
$Q=\{D,a_0\}, k=2, \gamma=3$	{A, C, D, E}
$Q=\{D,a_1\}, k=2, \gamma=3$	{D, I, F}
$Q=\{D,a_0\}, k=3, \gamma=3$	{ }
$Q=\{D,a_0,a_1\}, k=2, \gamma=5$	{A, D, E, F, I}
$Q=\{D,a_0,a_1\}, k=3, \gamma=5$	{ }

Proposed Methodology

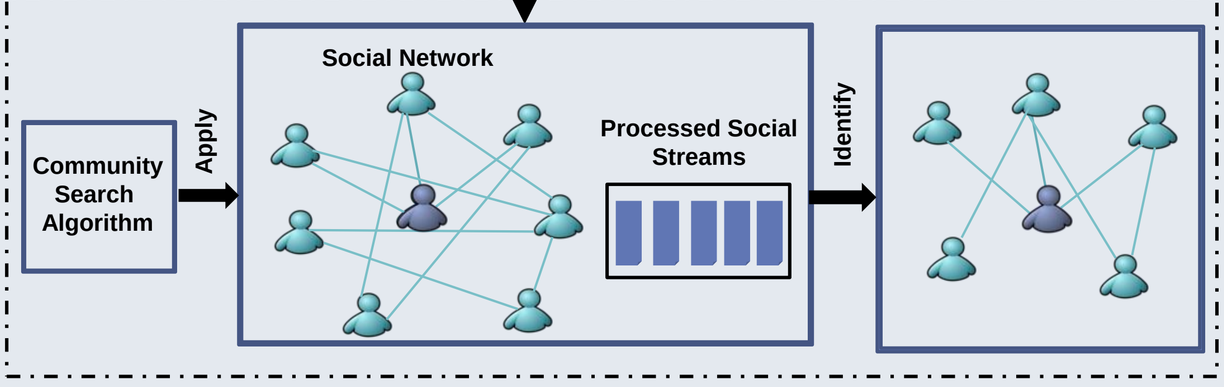
Stage 1



Stage 2



Stage 3



Experimentation and Finding

Activeness Score

The first factor f_1 is the likelihood that a user u_i performs an activity in Q.

$$f_1(u_i, \psi_{u_i}) = \frac{|ACTS(u_i, \psi_{u_i})|}{|ACTS(u_i, *)|} \quad (1)$$

The second factor f_2 is the participation of user u_i related to Q in comparison to the most active participant user related to Q in the network.

$$f_2(u_i, \psi_{u_i}) = \frac{|ACTS(u_i, \psi_{u_i})|}{max_{u_z \in U^Q} |ACTS(u_z, \psi_{u_z})|} \quad (2)$$

Then, the activeness (denoted as σ) of u related to Q is,

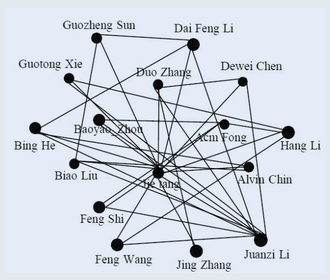
$$\lambda_{(u_i,\psi_{u_i})} = f_1(u_i, \psi_{u_i}) \times f_2(u_i, \psi_{u_i}) = \frac{|ACTS(u_i, \psi_{u_i})|^2}{|ACTS(u_i, *)| \times max_{u_z \in U^Q} |ACTS(u_z, \psi_{u_z})|} \quad (3)$$

$$\sigma_{(u_i,\psi_{u_i})} = \frac{\lambda_{(u_i,\psi_{u_i})}}{max_{u_z \in \mathcal{A}_Q}\{\lambda_{(u_z,\psi_{u_z})}\}} \quad (4)$$

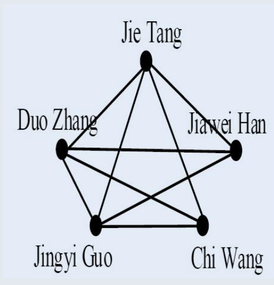
Date Sets

Dataset	# of Nodes	# of Nodes	# of activities
SNAP	400,000	5,357,560	573,832
Co-Author	327	1,180	807

Results

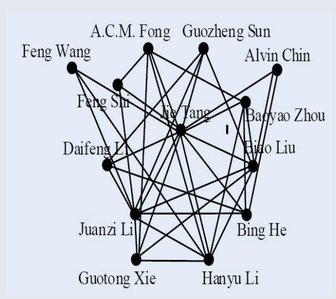


k=3, Q={ }

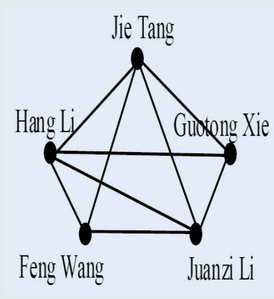


k=3, Q={AI}, r=1

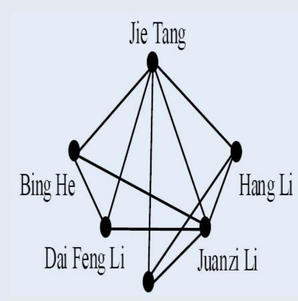
k=3, Q={AI,DM}, r=1



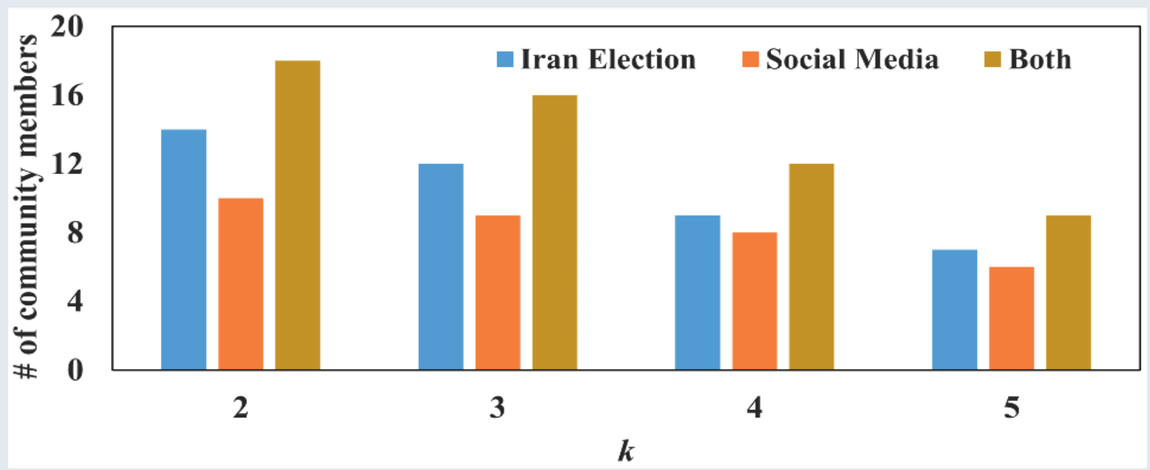
k=3, Q={AI, ML,DM}, r=1



k=3, Q={AI, ML}, r=2



k=3, Q={AI, ML, DM}, r=3



Community evolution(SNAP Dataset)

Conclusion and Future Work

This Work

-Measures an **activeness score** function for the candidate community members according to their common interests by the given query.

-Conducts experiment on two **real data sets** demonstrated the effectiveness of our proposed method.

Future Work

-To get more **recent activeness** score of the users' **time frame or dynamic network** can be included.

References

1. Huang, X., Lakshmanan, L. VS.: Attribute-driven community search. In: VLDB,pp. 949–960 (2017)
2. Zhou, Y., Cheng, H., Yu, J. X.: Graph clustering based on structural/attribute similarities. In: VLDB, pp. 718–729(2009).
3. Anwar, M. M., Liu, C., Li, J..: Discovering and tracking query oriented activeonline social groups in dynamic information network. In: WWWJ, pp. 1–36 (2018)
4. Fang, Y., Cheng, R., Luo, S., Hu, J.: Effective Community Search for Large Attributed Graphs. In: VLDB, pp. 1233–1244 (2016

NSysS 2018, Dec 18-20, 2018, Dhaka, Bangladesh