Lecture Notes: Social Networks: Models, Algorithms, and Applications
Lecture 2: Jan 19, 2012
Scribes: Preethi Ambati and Azar Aliyev

Example 3: Viral Marketing and the vaccination policy problem

Diffusion process on network has been studied for a long time. Here are a few well known examples:

Ryan and Gross "The diffusion of hybrid seed corn in two Iowa communities" [1].
Farmers typically hear about hybrid corn seeds from sales persons, but would usually buy them only because of interpersonal network influence. Therefore they found that mass communication functioned as the source of initial information, while interpersonal networks functioned as the influence over the farmers' decisions to adopt.

Coleman, Katz, Menzel Adoption of tetracycline by physicians in the U.S [2].

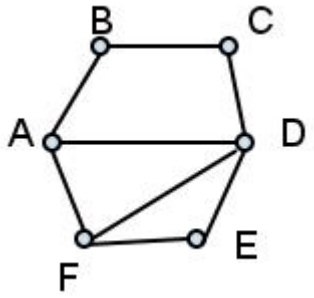
Domingos and Richardson asked the following question:
"If we can try to convince a subset of individuals to adopt a new product/innovation so as to trigger a 'cascade' of further adoptions, who should we target?"[3]

Kempe, Kleinberg, Tardos KDD Kempe, Kleinberg, Tardos suggested an algorithmic answer to the above question. [4]

Influence Maximization Problem (IMP)
Input: Input to the problem is a graph G = (V, E), where V is vertex set of graph and E is edge set of graph and a positive integer K.
Output: Output to the problem is a size K subset A $\subseteq$ V with maximum influence.
To make this problem precise, we need a definition of "influence".
To define what "influence" means, we need a model for how diffusion occurs.
Kempe et. al consider several diffusion models including the Linear Threshold Model and the Independent Cascade Model, etc.

Independent Cascade Model.

1. Initially some subset of A$\subseteq$ V of the vertices are active. The rest are inactive.
2. Diffusion proceeds in discrete steps.
3. When vertex v becomes active (in a particular time step), it has one chance to activate each neighbor w that is currently inactive. W becomes active due to v with probability $P_{v,k}$
4. If w is activated in time step t, it has the opportunity to activate neighbors in time step t+1.



In time step t=0,

A has the opportunity to activate its neighbors.

B is activated with some probability via A.

D is activated with some probability via A and with some probability via E.

F is activated with some probability via A and with some probability via E.

5. The process ends in a time step in which there are no activations.

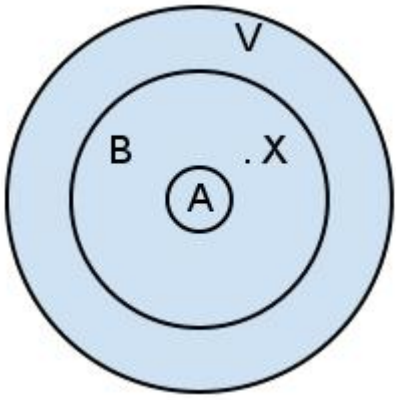
In the next step, t=1
A and E are no longer a factor because they have already activated their respective neighbors.

This alghoritm can be applied in a disease spread model (SI model where S stands for susceptible and I for infected).

Suppose we choose A $\subseteq$ V as our set of "initiators" and start the independent cascade model (ICM).
Either path can be taken with some probability.
$\sigma(A)$ = expected number of vertices which get activated if we start ICM with initial set A of active nodes.
We want to choose that subset which maximizes $\sigma(A)$
Kempe, Kleinberg and Tardos (KKT) biggest contribution:
For many diffusion models including LTM and ICM σ satisfies very nice property called sub modularity. Note that σ is a set function. i.e. σ assigns a number to each subset A$\subseteq$ V
$\sigma : 2^v$ Notation for "set of all subsets of V".

$$\sigma(A \cup \{X\} - \sigma(A) \geq \sigma(B \cup \{X\}) - \sigma(B)$$

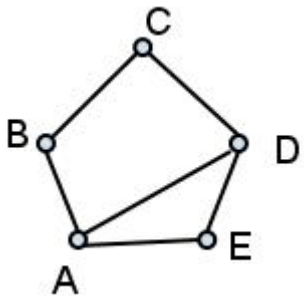
The above represents "diminishing returns" (Discrete property of diminishing returns).
Consider A, B as arbitrary sets such that B is superset of A.



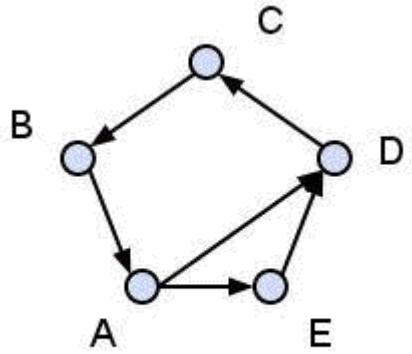
Graph – Theoretical Terminology:

O: nodes
/: edges
{D, E}: undirected edges
(D, E): directed edges

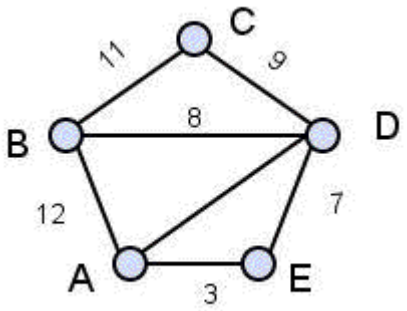
Undirected unweighted graph:



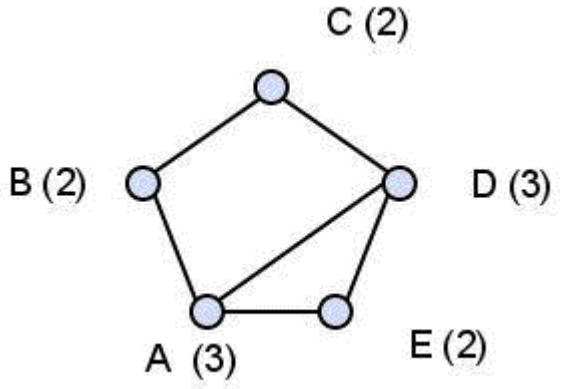
Directed weighted graph:



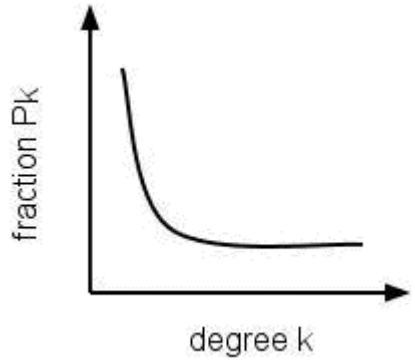
Directed weighted Graph:



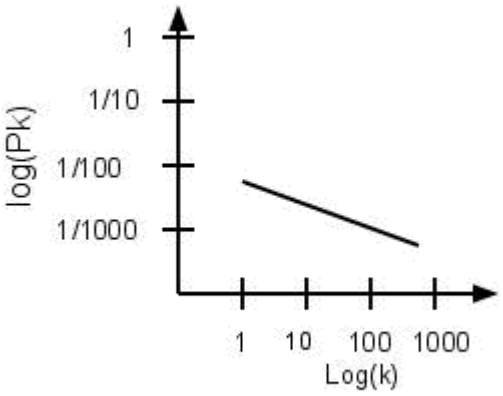
The notions degree distribution and clustering show up a lot in the context of social network.



Neighbors of D: A, C, E
N(V) : set of neighbors of V
Degree (v): size of set of neighbors of v = |N(v)|
The degree of distribution of a graph is the sequence P0, P1, P2,.....
PK = fraction of vertices with degree K.
P0 = 0/5
P1 = 0/5
P2 = 3/5
P3 = 2/5
Pk = 0 ? K > 3
If we pick vertex $v \in V$ uniformly at random from G the probability $Prob\ [\ deg\ (v)\ :\ k]\ = P_k$.
Famous example of a degree distribution is power law degree distribution.
Faloutsos, Faloutsos, Faloutsos Paper appeared in the ACM SIGCOMM 1999.
"The internet graph has power-law degree distribution"
In the internet graph, the vertices were the routers and the edges were one hop ip connections between pairs of routers
The graph was plotted with degree K on X-axis and fraction P_k along the Y-axis. The result was as below:



Then the scale was changed and the graph was as below.



The observation was that the distribution was linear. This means that log P_k can be written as a linear function.
Log P_k = - α . log (k) + c
Where $\alpha > 0$,
α, c are arbitrary constants
Raising both sides to the power of 10, we get
$P_k = 10^{(} - \alpha \; log \; (k)) \; 10 \; c \;\; = \;\; 10 \; c \; * \; K \; -- \; \alpha \;\;\; = \; C. \; K \; -- \; \alpha$Where C = 10 c
P(K) = C. K – α This is the power law distribution.
Consider a www graph - directed graph. In www graph if you take the in-degrees separately to out-degrees, the graph satisfies the power law distribution in both cases.
Apart from craters on the moon and sizes of the earth quakes, a lots of diverse phenomena satisfy the power law of distribution. People were excited that man-made artifacts also satisfied the power law distribution.
This led to a huge explosion of research in network science on a power law distribution.
To read more about this controversy, please refer to
- Lakhina et al. INFOCOM 2003, “Sampling biasis in IP Topology Measurements”
- Achlioptas et al STOC 2005

References

[1] Ryan and Gross. (1943), "The Iowa Study of Hybrid Seed Corn: The Adoption of Innovation." Rural Sociology, 8(1): 15-24.

[2] Coleman, Katz, Menzel. (1966), Medical Innovation: A Diffusion Study, Inidianapolis, The Bobbs-Merril Company, Inc.

[3] Domingos, P., Richardson, M (2001): Mining the Network Value of Customers. In: ICDM

[4] D. Kempe, J. Kleinberg, E. Tardos.(2003) Maximizing the Spread of Influence through a Social Network. Proc. 9th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining,