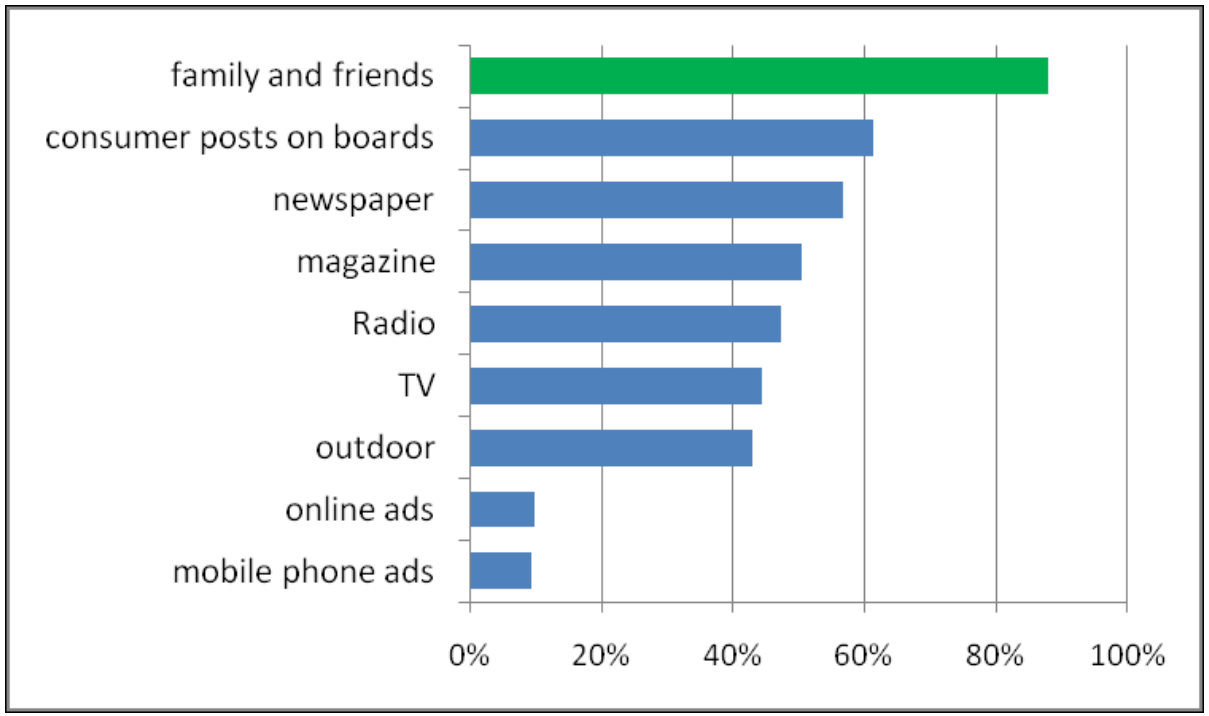
Viral Marketing & Influence Maximization

Word-of-mouth

- Social network plays a fundamental role as a medium for the spread of influence among its members.

Level of trust on different types of ads*



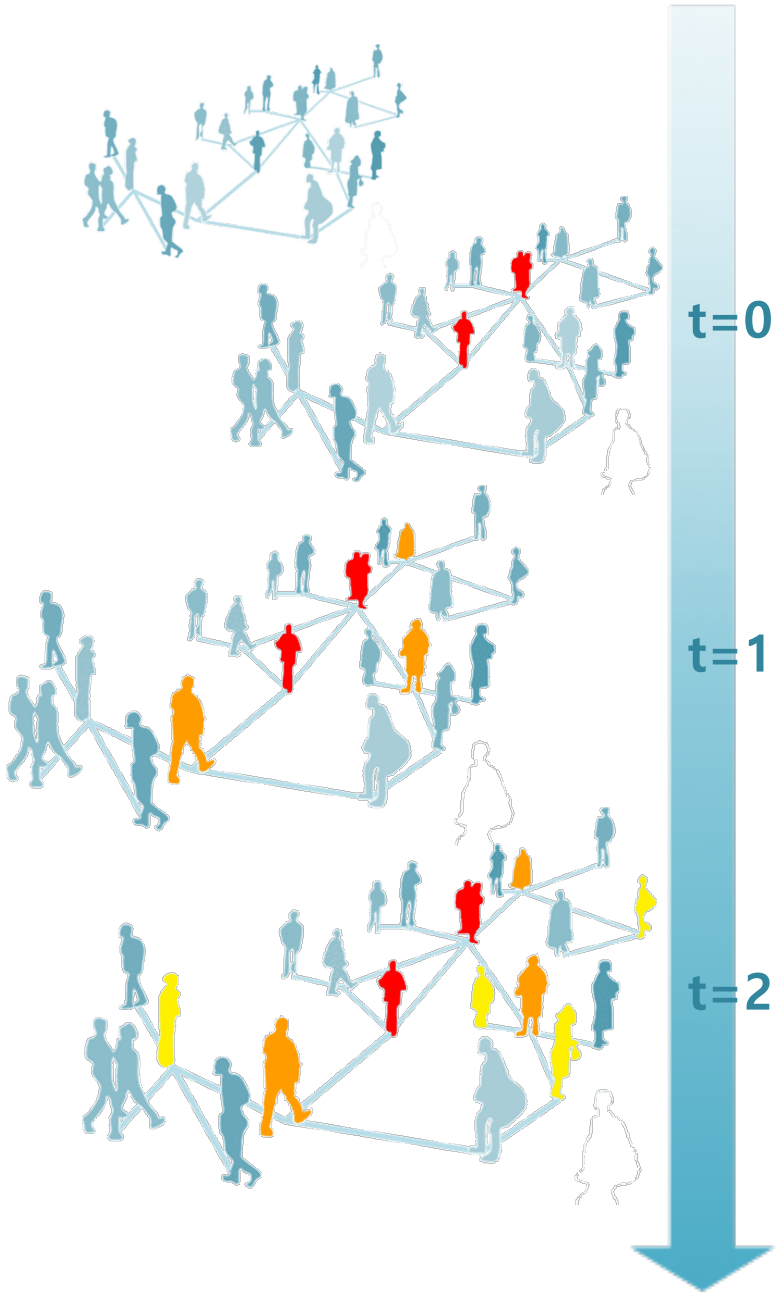
*source from Forrester Research and Intelliseek

"발 없는 말이 천리 간다."

Lanaa

Viral Marketing

- Interests
  - **Budget**: Who?
  - **Influence**: How many?
  - **Diffusion Efficiency (or marketing efficiency)**
    - Diffusion Time: How long?
    - Diffusion Number: How many?
- Influence Maximization
  - Given graph G (V,E), select “k” vertices to maximize (expected) diffusion efficiency



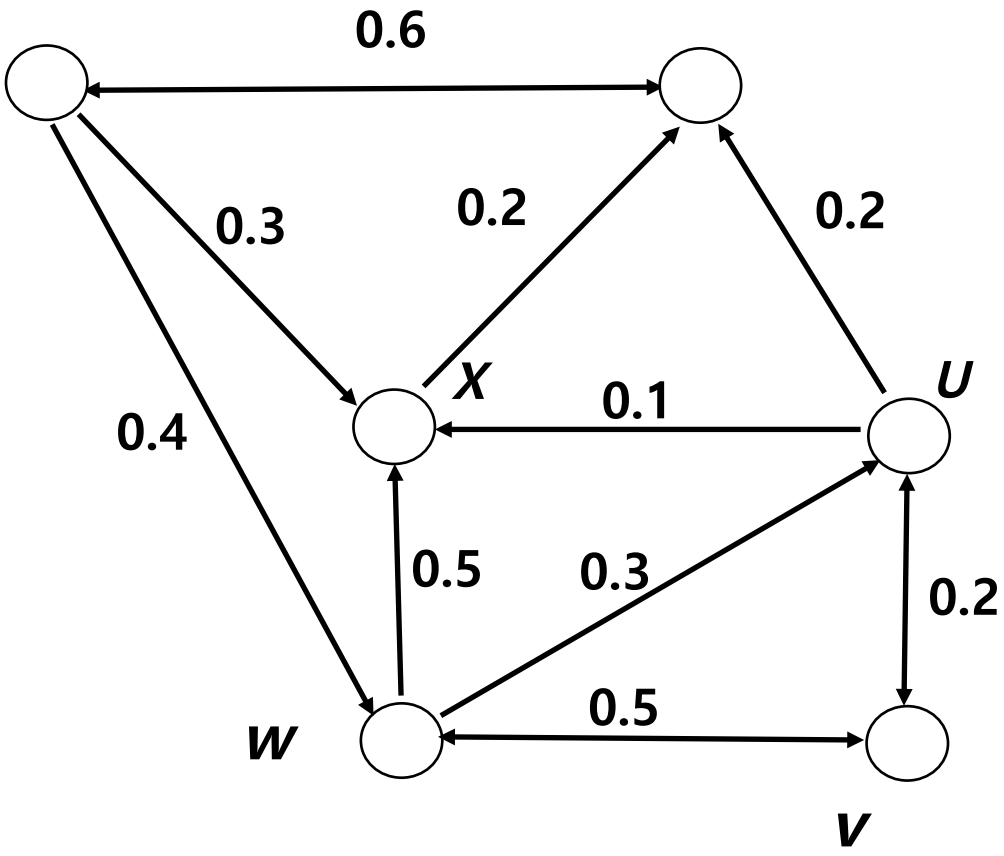
Lanaa

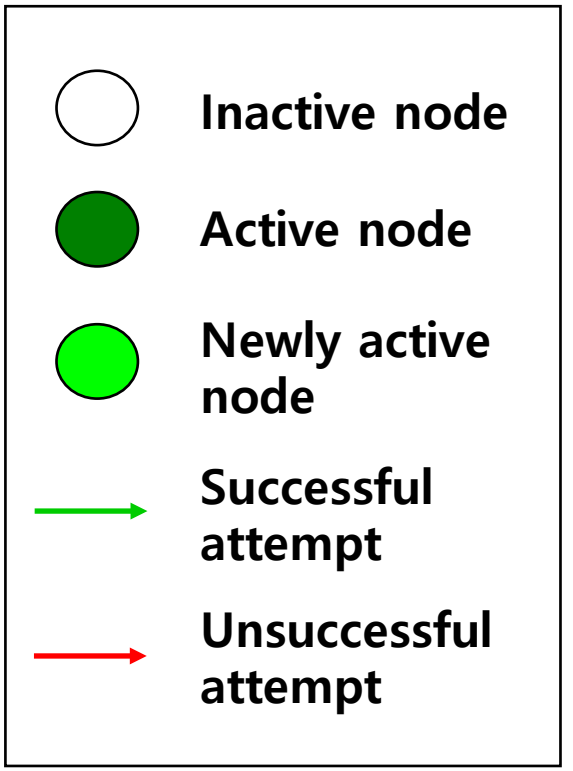
Diffusion Model
&
Objective

Independent Cascade Model

- When node v becomes active, it has a **single** chance of activating each currently inactive neighbor w.
- The activation attempt succeeds with probability p_{vw}.

Example





Stop!

Independent Cascade Model

- Let $U(C)$ denote the corresponding (random) set of ultimately activated nodes by active seed set C.
- $U(C)$ is a random variable thus we want to maximize $F(C) = E(|U(C)|)$ subject to $|C| < k$.

IF-MAX Problem

$$\begin{aligned} &Maximize\ F(C) := E(|U(C)|) \\ &Subject\ to\ |C| < k \end{aligned}$$

Problem Hardness
&
Efficient Approximation

Solving IF_MAX is NP-hard

- **Theorem NP-hard.**
  IF_MAX is NP-hard for the Independent Cascade model.
  - (sketch of proof)

    A simplified version of influence maximization problem in IC model is equivalent to NP-hard *Set Cover* problem. Thus, the original influence maximization problem is NP-hard problem as its simplified version.

    The simplified version of influence maximization problem:
    - Consider seed set C with cardinality k.
    - p_{ij} is 1 if i and j are connected. (or consider a sample path)
    - Define C_i as coverage by node i in C.