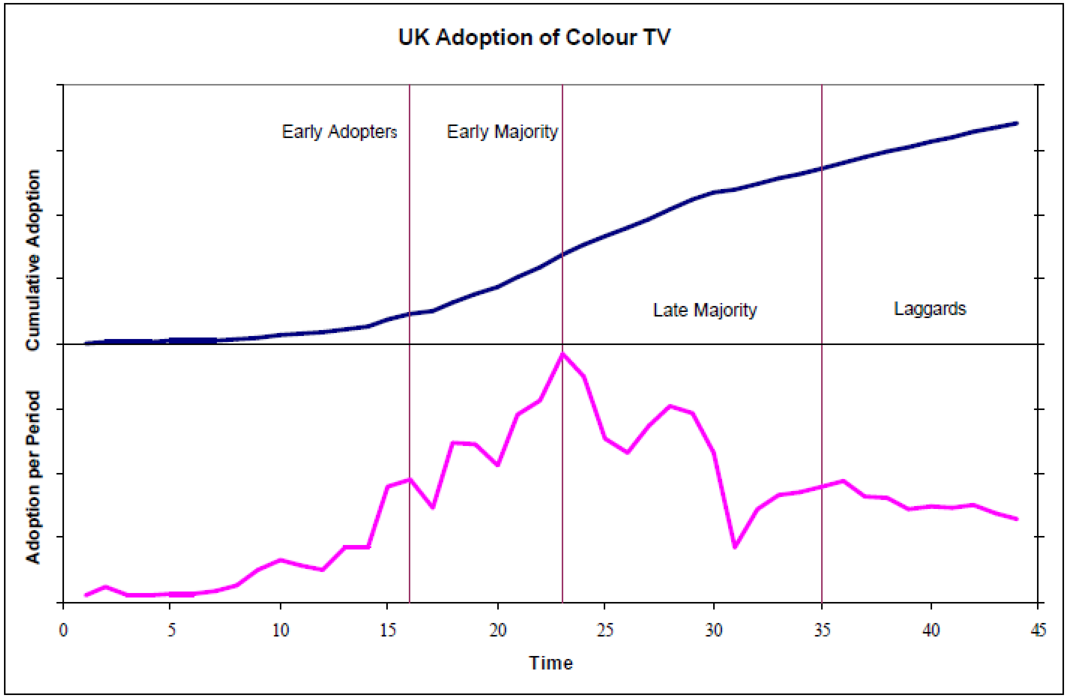
Modeling Diffusion in Social Networks using Network Properties

Minh-Duc LUU*, Ee-Peng LIM,

Tuan-Anh HOANG, Freddy Chong Tat CHUA

Why Diffusion in Social Networks?

Diffusion of items (videos, news, photos, etc) is important and ubiquitous in social networks



Proper models of diffusion can predict:

- Rate of adoption at a particular time
- The time of peak demand
- The magnitude of peak demand

Applications of Diffusion Models in Telecommunications, Nigel Meade
http://www.itu.int/ITU-D/finance/work-cost-tariffs/events/expert-dialogues/forecasting/meade-presentation.pdf

Why Modeling Diffusion using Network Properties?

For item diffusion we have micro and macro models.

	Micro models	**Macro** models
Work at	**Individual** level	**Network** level
Representatives	Independent Cascade (**IC**)[1], Linear Threshold (**LT**)[2]	**Bass Model**[3] and its extensions
Parameters	**Local**, each user has his own parameters	**Global**, for the whole network
Network properties	Exploit (**+**)	Do NOT exploit (**-**)
No. of parameters	So many (**-**)	Just a few (**+**)

[1] Goldenberg et al. (2001)*Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth*

[2] Granovetter, M. (1978) *Threshold Models of Collective Behavior*

[3] Bass, F. M. (1969) *A new Product Growth Model for Consumer Durables*

Research Questions

How can macro models exploit network properties (e.g. degree distribution)?

In this work:

Q1) How to model diffusion in a network given its degree distribution?

Q2) How to combine parameters of diffusion and of degree distribution to give a better model?

Concepts & Notations

- *N:* network size.
- *f(t)*: instantaneous fraction of adopters at time *t*
- *F(t)*: cumulative fraction of adopters at time *t*

$$f(t) = F'(t)$$

- a_t : adoption at *t* , A_t : adoption before *t*.

Observe that:

$$\underbrace{f(t)}_{\substack{\text{fract of}\\ \text{new adopters}}} = \underbrace{(1 - F(t))}_{\text{nonadopters at } t} \cdot \underbrace{\mathrm{E}\left[P\left(a_t \mid \overline{A_t}\right)\right]}_{\text{avg. adopting prob}}$$

Ordinary differential equation (ODE) linking *F(.), P(.)*:

$$F'(t) = (1 - F(t)) \cdot P\left(a_t \mid \overline{A_t}\right)$$

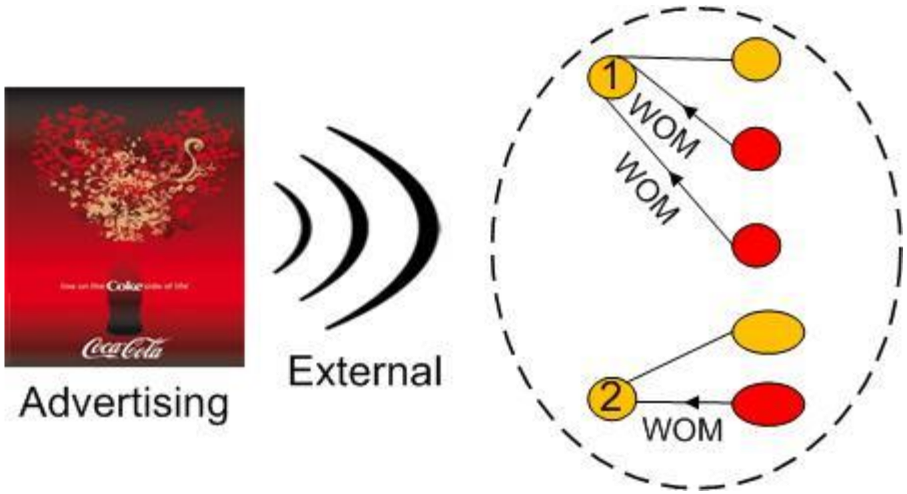
Goal: estimate the adoption probability *P(.)* as a function of *F(.)*

SMU SINGAPORE MANAGEMENT UNIVERSITY

mda Media Development Authority Singapore

Carnegie Mellon

Estimation of Adoption Probability General Case



Internal comes from WOM (word of mouth)

Contributions from external & internal influences are weighted with w_e and $1\text{-}w_e$ respectively.

$$P\left(a_t \mid \overline{A_t}\right) = w_e \cdot P_{ext}\left(a_t \mid \overline{A_t}\right) + \left(1 - w_e\right) \cdot P_{int}\left(a_t \mid \overline{A_t}\right)$$
$$= w_e \cdot p_e + \left(1 - w_e\right) \cdot P_{int}\left(a_t \mid \overline{A_t}\right)$$

SMU SINGAPORE MANAGEMENT UNIVERSITY

mda Media Development Authority Singapore

Carnegie Mellon

Bass Model (BM)

Assumptions of BM:

B1) Each user can influence every other user.

$$|u\text{'s adopted neighbors}| = N \cdot F(t), \quad \forall u, \forall t$$

B2) Internal influence is proportional to No. of adopted neighbors: $P_{int}\left(a_t \mid \overline{A_t}\right) = q_1 \cdot N \cdot F(t)$

$$\Rightarrow P\left(a_t \mid \overline{A_t}\right) = p + q \cdot F(t) \qquad (*)$$

where $p = w_e \cdot p_e$ and $q = (1 - w_e) \cdot q_1 \cdot N$

(*) combines with the ODE
→ Bass Model (1969)

$$F(t) = \frac{e^{[(p+q)t]} - 1}{e^{[(p+q)t]} + q/p}$$

$$f(t) = F'(t) = \frac{(p+q)^2}{p} \cdot \frac{e^{[(p+q)t]}}{\left\{e^{[(p+q)t]} + q/p\right\}^2}$$

SMU SINGAPORE MANAGEMENT UNIVERSITY | mda Media Development Authority Singapore | Carnegie Mellon

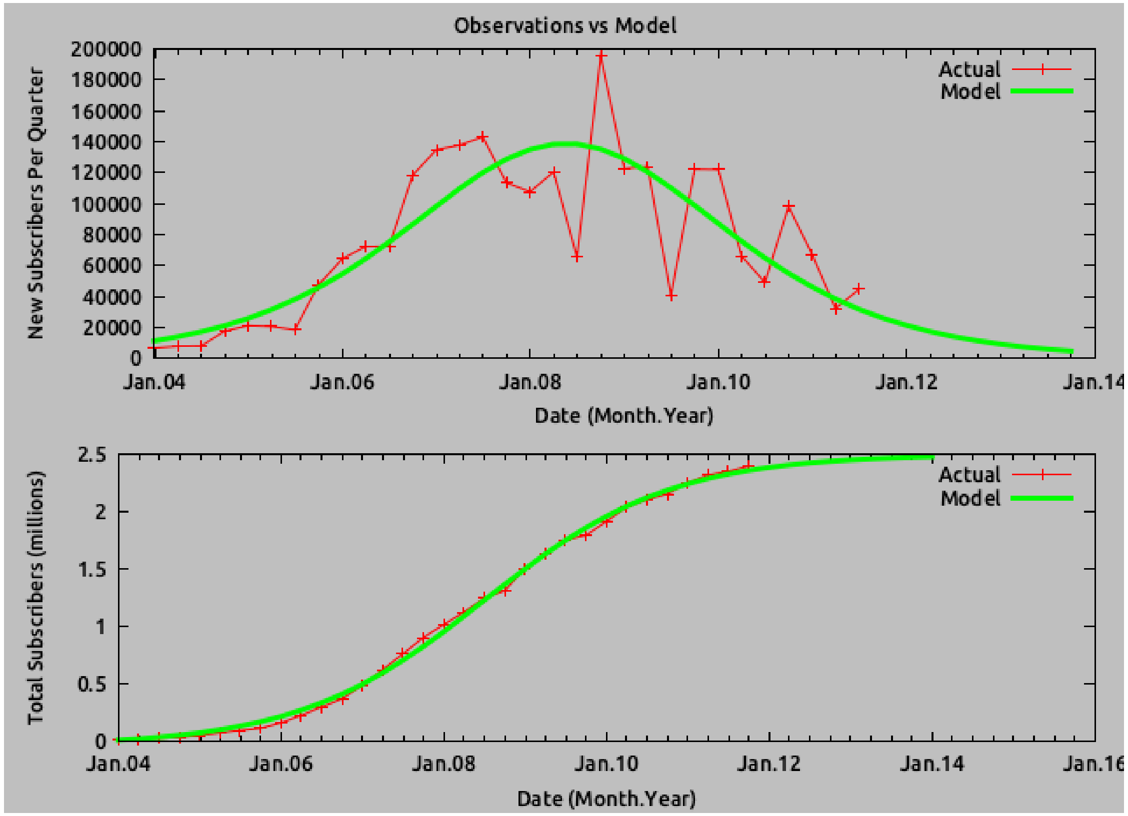
Bass Model (cont.)

"…Bass model ignores the network structure…"
[Xiaodan S. et al, WWW 07]

B1) Each user can influence/influenced every other user

Each user can influence/influenced only his friends
→his adoption prob. depends on his degree

Bass Model (cont.)



http://serial.innovatiasystems.eu/cms/

Adoption Probability for Specific Degree Distributions

- Given any degree distribution *P(k)*, we obtained the formula:

(**Internal** adoption probability)

$$P_{\text{int}}\left(a_t \mid \overline{A_t}\right) = \sum_{k=1}^{N-1} P(k) \sum_{j=0}^{k} \left[\binom{k}{j} F_t^{\,j} \left(1 - F_t\right)^{k-j} P\left(a_t \mid \overline{A_t}, j\right) \right]$$

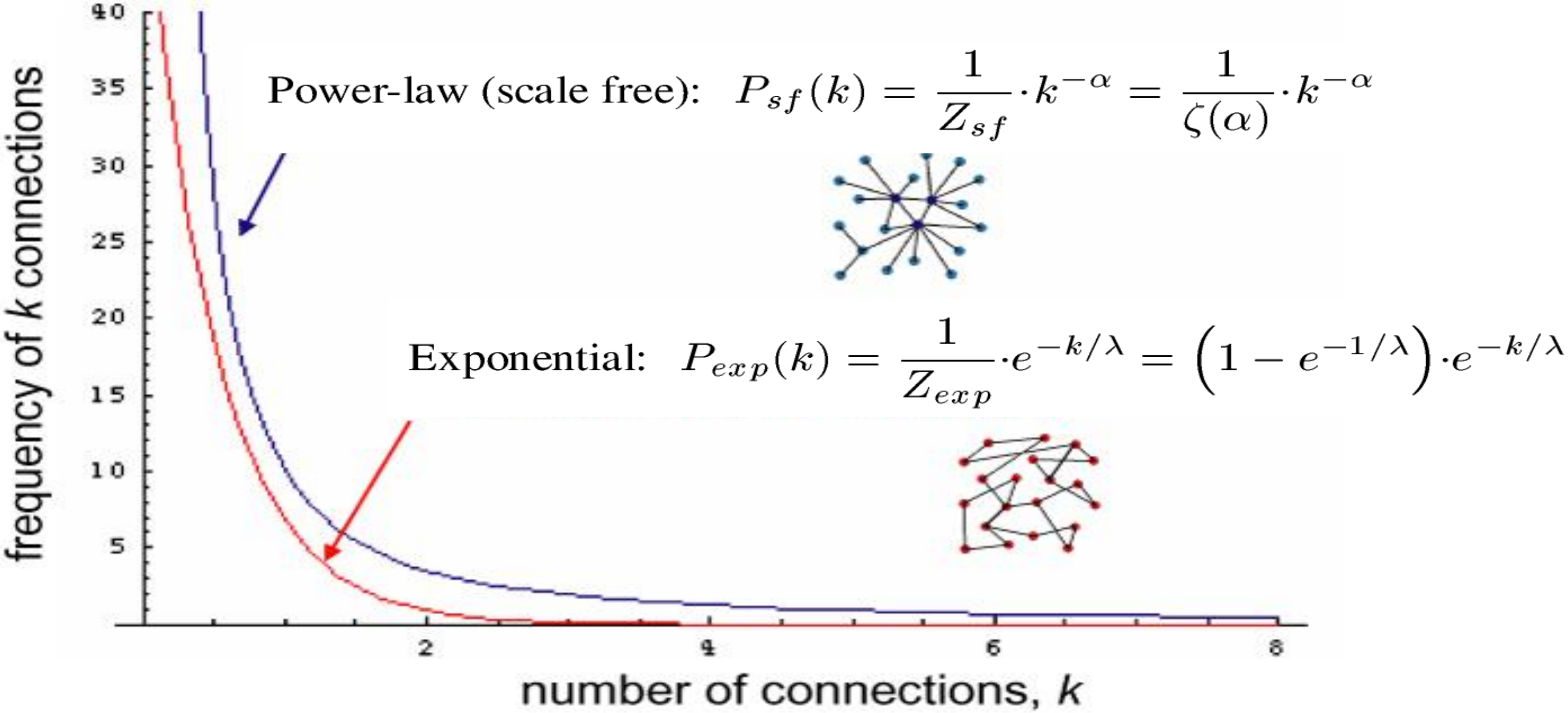
where $F_t \equiv F(t)$ and $P\left(a_t \mid \overline{A_t}, j\right)$ is the prob of adopting given that a user has j adopted neighbors

- Still keep B2), linear influence: $P\left(a_t \mid \overline{A_t}, j\right) = c \cdot j$

where c is a constant

To complete estimation, needs specific degree distributions!

Specific Degree Distributions



Pagel *et al. BMC Evolutionary Biology*
2007 **7**(Suppl 1):S16

Parameter of degree distribution: α (power law) or λ (exponential)

Estimation of Internal Adoption Probability

Linear assumption & specific degree distribution provide estimations:

1) Scale free network:

$$P_{int}^{sf}(a_t|\overline{A_t}) = \frac{\zeta(\alpha-1)}{\zeta(\alpha)} \cdot c \cdot F_t$$

where $\zeta(\alpha) = \sum_{k=1}^{\infty} k^{-\alpha}$ is the Riemann Zeta function

2) Exponential network:

$$P_{int}^{exp}(a_t|\overline{A_t}) = \frac{e^{-1/\lambda}}{1-e^{-1/\lambda}} \cdot c \cdot F_t$$

These estimations → two models in our work.

SMU SINGAPORE MANAGEMENT UNIVERSITY

mda Media Development Authority Singapore

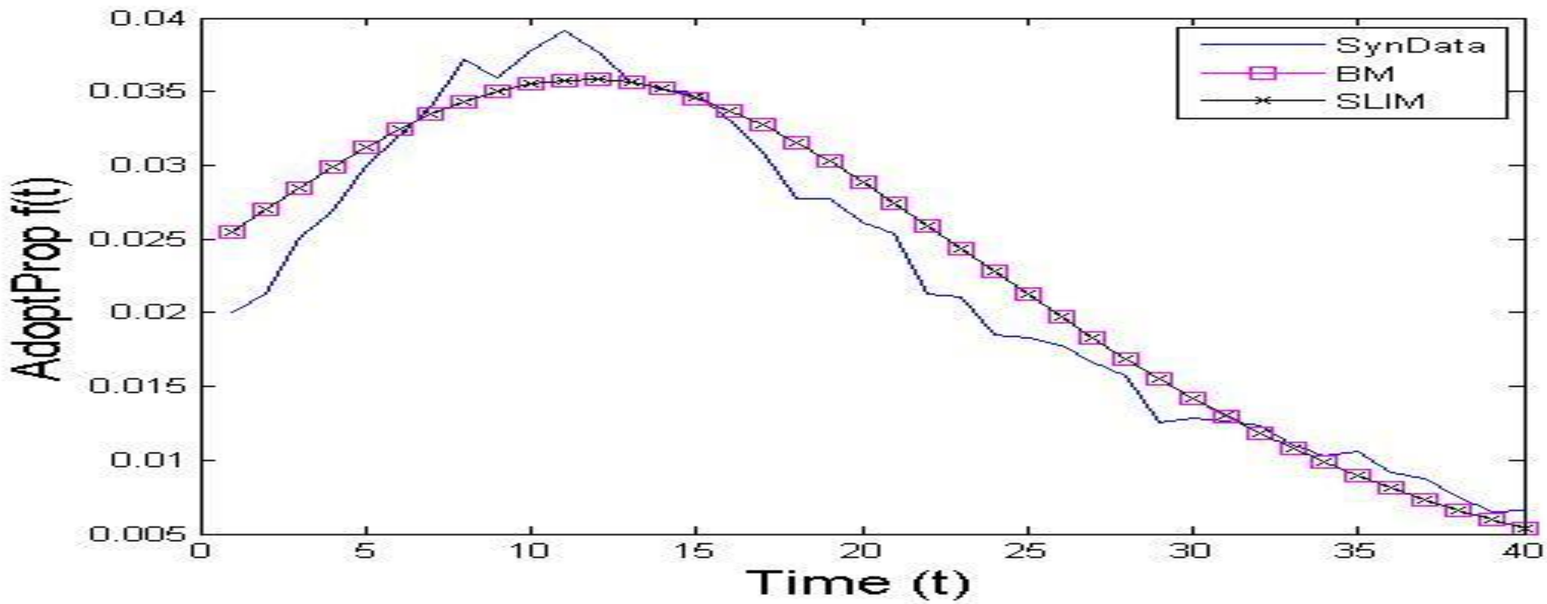
Carnegie Mellon

Proposed Models

1. SLIM (Scale-free Linear Influence Model): Scale-free network.
2. ELIM (Exponential Linear Influence Model): Exponential network

Remarks:

- Give more rigorous estimation of adoption probability by combining parameters of diffusion and of degree distribution.
- Give the **same** fitting error as BM though!



SMU SINGAPORE MANAGEMENT UNIVERSITY

mda Media Development Authority Singapore

Carnegie Mellon

What is the problem?

$$\underbrace{f(t)}_{\text{fract of new adopters}} = \underbrace{(1-F(t))}_{\text{nonadopters at } t} \cdot \underbrace{\mathrm{E}\left[P\left(a_t \mid \overline{A_t}\right)\right]}_{\text{avg. adopting prob}}$$

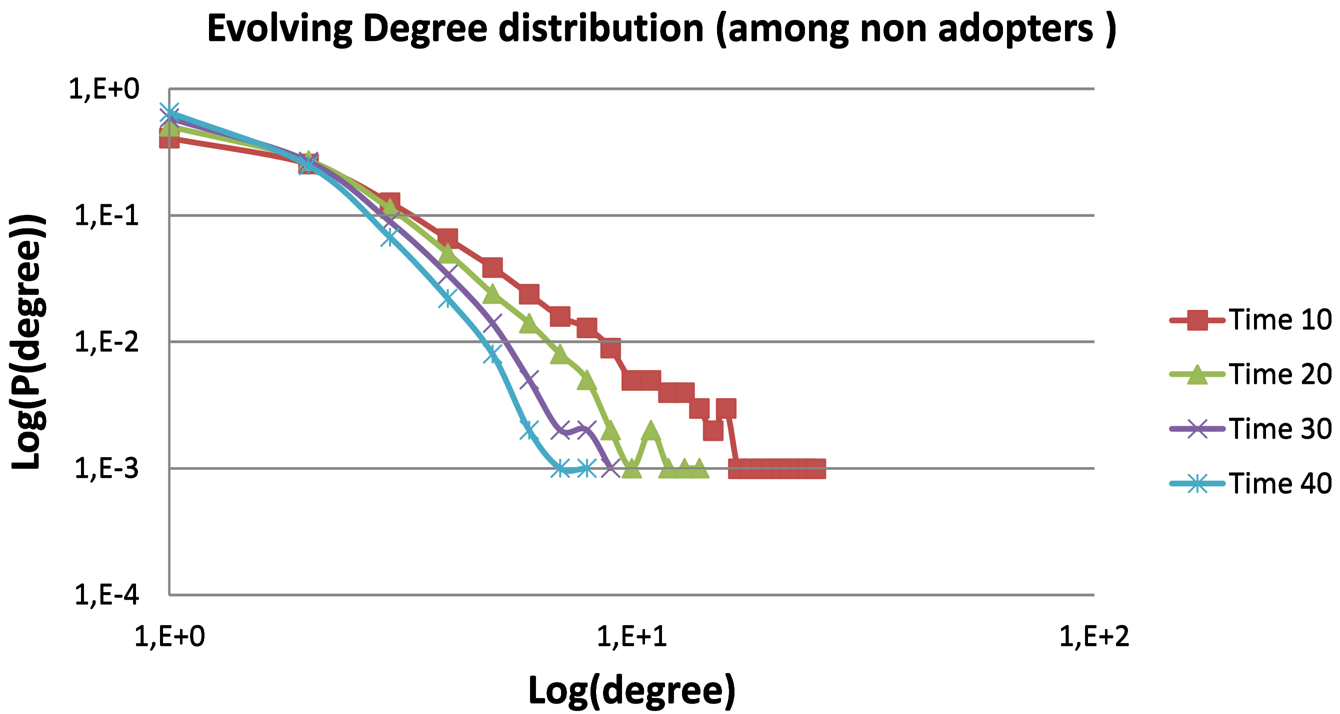
- Is it correct to use degree distribution of the whole network for *P(k)*?

  NO. Should use degree distribution over the set of non-adopters (NA).

- NA changes over time → its degree distribution also changes?

  YES.

Degree Distribution is Dynamic !!



Synthetic scale-free network; 27,289 nodes and 27,031 edges (α_0=3).

As time proceed, users with high/low degs are more/less likely to adopt and leave/stay NA set. Thus later distributions are more biased to low degrees.

SMU SINGAPORE MANAGEMENT UNIVERSITY

mda Media Development Authority Singapore

Carnegie Mellon

Multi-Stage Model (MLIM)

For different time pts, need to employ different models.
How to decide the proper model?

Heuristic approach: in a **short** duration, degree distribution does NOT significantly change.

- Divide diffusion process into *n* stages. Each has short duration (< 10 time steps).
- For each stage, choose between SLIM and ELIM the one that gives smaller fitting error.

→ Multi-Stage Model

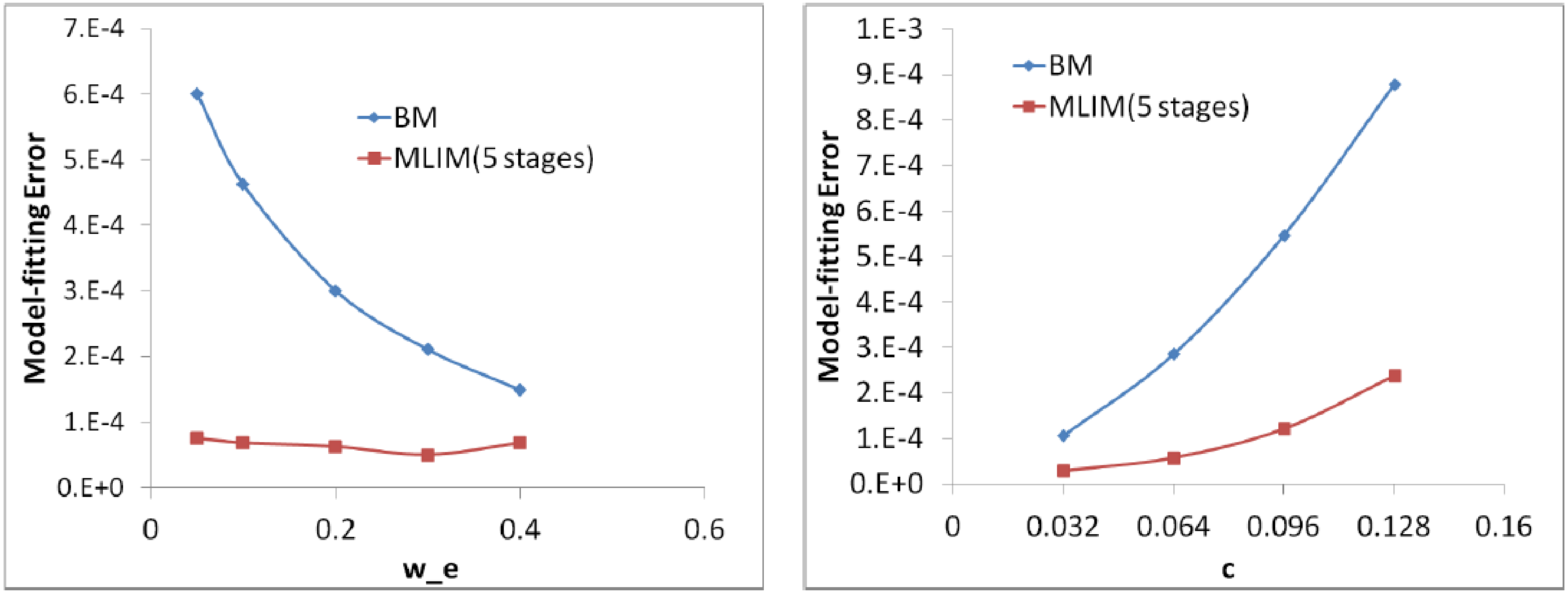
SMU SINGAPORE MANAGEMENT UNIVERSITY

mda Media Development Authority Singapore

Carnegie Mellon

Experiments on Synthetic Data

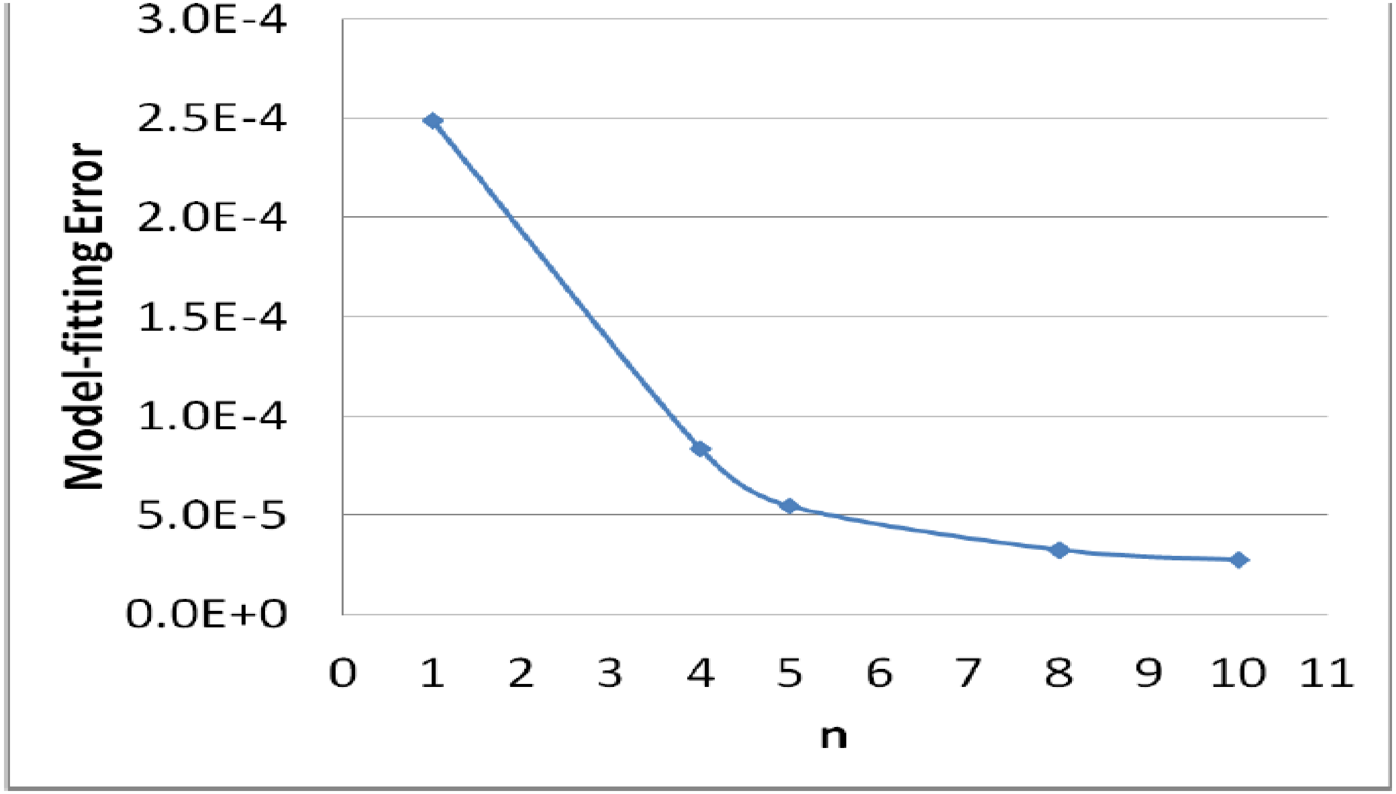
- Network: 28,172 nodes; 34,578 edges (α=2.5).
- Evaluation metrics: model-fitting error (LSE) & parameter-learning error



(a) w_e in $\{0.05, 0.1, 0.2, 0.3, 0.4\}$ (b) c in $\{0.032, 0.064, 0.096, 0.12\}$

SMU SINGAPORE MANAGEMENT UNIVERSITY
mda Media Development Authority Singapore
Carnegie Mellon

Experiments on Synthetic Data

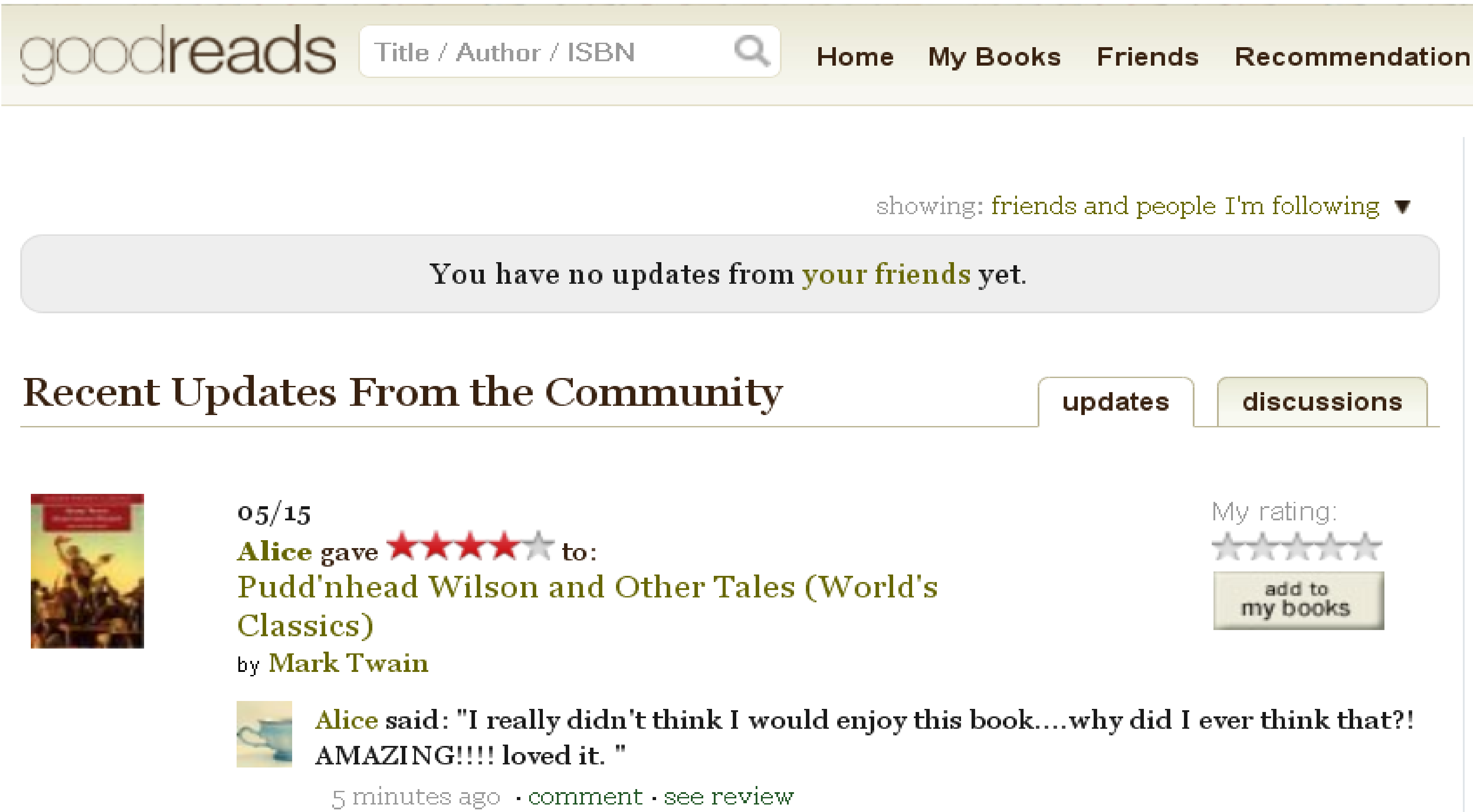


(c) $n \in \{1, 4, 5, 8, 10\}$

n=1 corresponds to BM

Real-world Dataset

From Goodreads network (www.goodreads.com), ≅ 87K users; 159,442 follow links



SMU SINGAPORE MANAGEMENT UNIVERSITY

mda Media Development Authority Singapore

Carnegie Mellon

Experiment Design

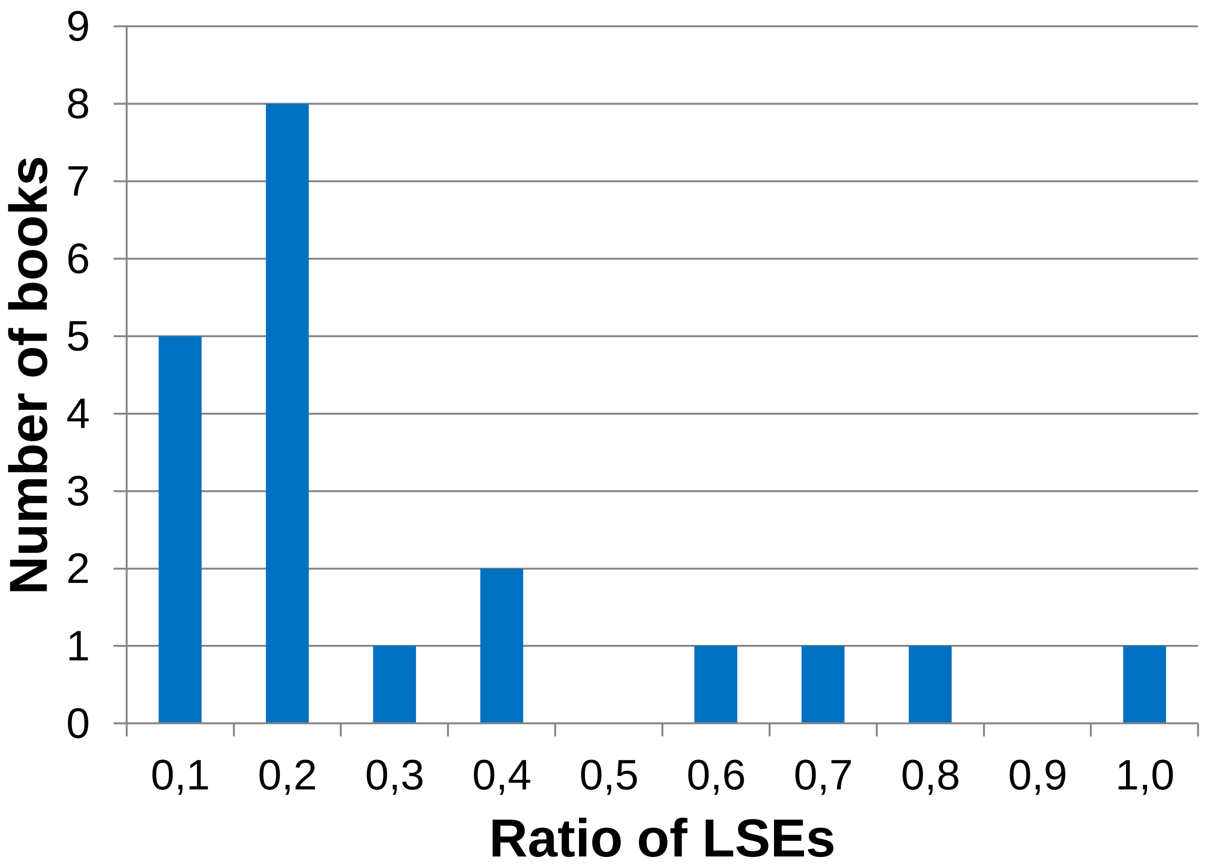
- Adopting a book ≅ writing review on it.
- Review data was collected for 73 popular books.
- Period: 05/2007 to 02/2011 (45 months).
- Filter out books with review data spans < 30 months
  → 20 books remain (Harry Potter 7, Breaking Dawn, …).

- Evaluation metric: ratio of model-fitting errors (LSE of MLIM over LSE of BM)
  → less than 1 shows improvement of our model.

Special thanks to Agus and Anh.T.H

SMU SINGAPORE MANAGEMENT UNIVERSITY · mda Media Development Authority Singapore · Carnegie Mellon

Results for Top-20 Popular Books



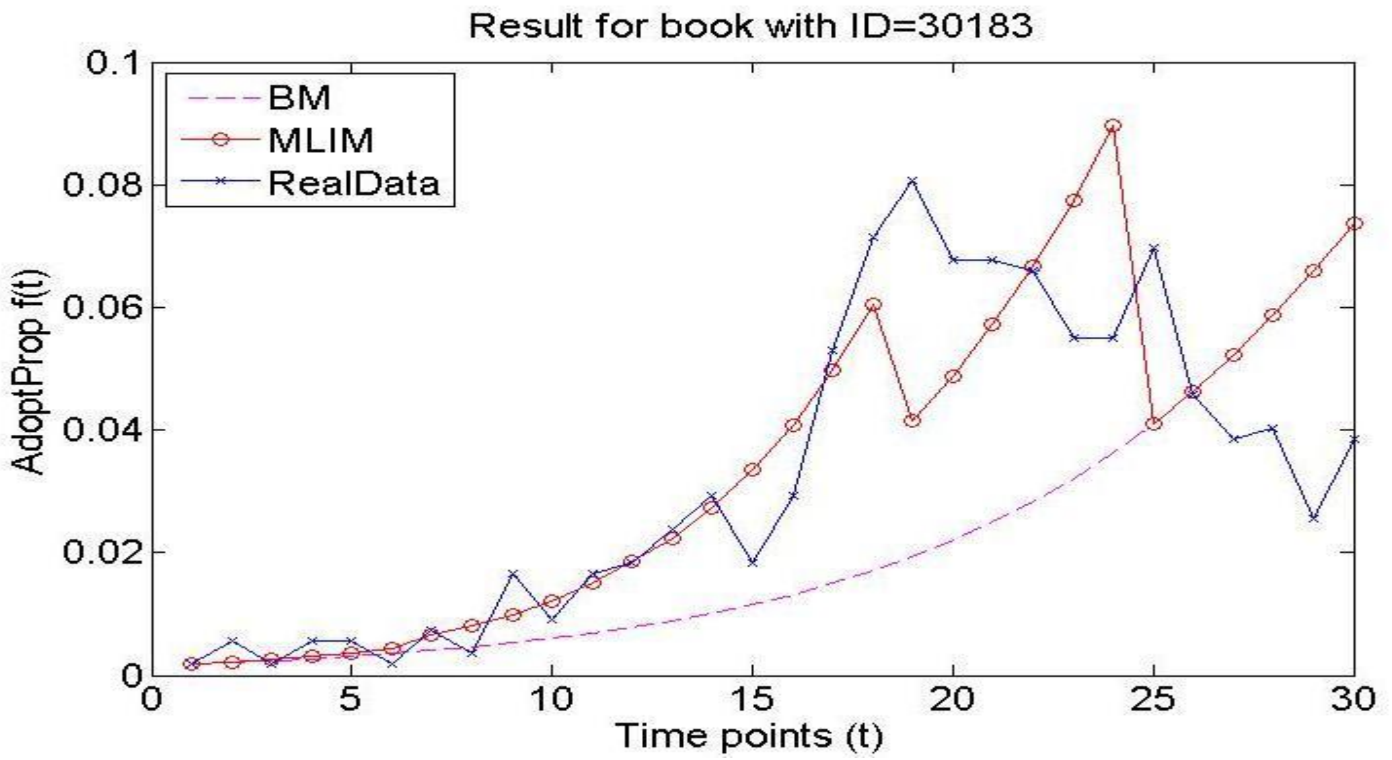
- MLIM outperforms BM for all top-20 popular books.
- 75% of books have error ratios less than ½.

SMU SINGAPORE MANAGEMENT UNIVERSITY

mda Media Development Authority Singapore

Carnegie Mellon

Zoom-in for One Book



Fitting result for *City of Ashes* by Cassandra Clare

MLIM provides significant improvement over BM in terms of fitting data.

Conclusion

- This work:
  - Proposed two models SLIM, ELIM for diffusion in scale-free and exponential networks respectively.
  - Proposed multi-stage model (MLIM) to deal with dynamic degree distribution.
- Future works:
  - Derive a more rigorous way to deal with dynamic degree distribution.
  - Replace linear influence by other (e.g. **quadratic, exponential**) influence?
  - Examine the effect of other network quantities on diffusion.

SMU SINGAPORE MANAGEMENT UNIVERSITY

mda Media Development Authority Singapore

Carnegie Mellon

SMU SINGAPORE MANAGEMENT UNIVERSITY

mda Media Development Authority Singapore

Carnegie Mellon

More Formulae

Adoption prob. for scale free and exponential network

$$P_{sf}\left(a_t \mid \overline{A_t}\right) = w_e \cdot p_e + \left(1 - w_e\right) \cdot \frac{\zeta(\alpha - 1)}{\zeta(\alpha)} \cdot c \cdot F_t$$

$$P_{exp}\left(a_t \mid \overline{A_t}\right) = w_e \cdot p_e + \left(1 - w_e\right) \cdot \frac{e^{-1/\lambda}}{1 - e^{-1/\lambda}} \cdot c \cdot F_t$$

SMU SINGAPORE MANAGEMENT UNIVERSITY

mda Media Development Authority Singapore

Carnegie Mellon

Formulae of SLIM, ELIM

$$F_{SLIM}(t) = \frac{\exp\left[\left(p + q_{SLIM}\right)\cdot t\right] - 1}{\exp\left[\left(p + q_{SLIM}\right)\cdot t\right] + \left(q_{SLIM} / p\right)}$$

where $p = p_e \cdot w_e$ and $q_{SLIM} = \left(1 - w_e\right) \cdot \frac{\zeta(\alpha - 1)}{\zeta(\alpha)} \cdot c$

$$F_{ELIM}(t) = \frac{\exp\left[\left(p + q_{ELIM}\right)\cdot t\right] - 1}{\exp\left[\left(p + q_{ELIM}\right)\cdot t\right] + \left(q_{ELIM} / p\right)}$$

where $p = p_e \cdot w_e$ and $q_{ELIM} = \left(1 - w_e\right) \cdot \frac{e^{-1/\lambda}}{1 - e^{-1/\lambda}} \cdot c$

SMU SINGAPORE MANAGEMENT UNIVERSITY

mda Media Development Authority Singapore

Carnegie Mellon