# REVIEW ON ZIEGLER'S

## *Towards Decentralized Recommender Systems*

by Richard Kwo
richardkwo@gmail.com

# About Me

郭方健 *(Richard Kwo)*

- 2006~2009 *Chengdu No.7 High School*
- 2009~present *Yingcai Experimental School, UESTC*

Interests

- *Science*
- *Reading*
- *Music, including rock, folk etc.*
- *Using, playing & programming GNU Linux/open-source stuff etc.*
- *Graphic design & Typography*

Contact me


- *richardkwo@gmail.com*
- *13882042045*

# About Cai-Nicolas Ziegler

IEEE Computer Society *Best Paper Award* for

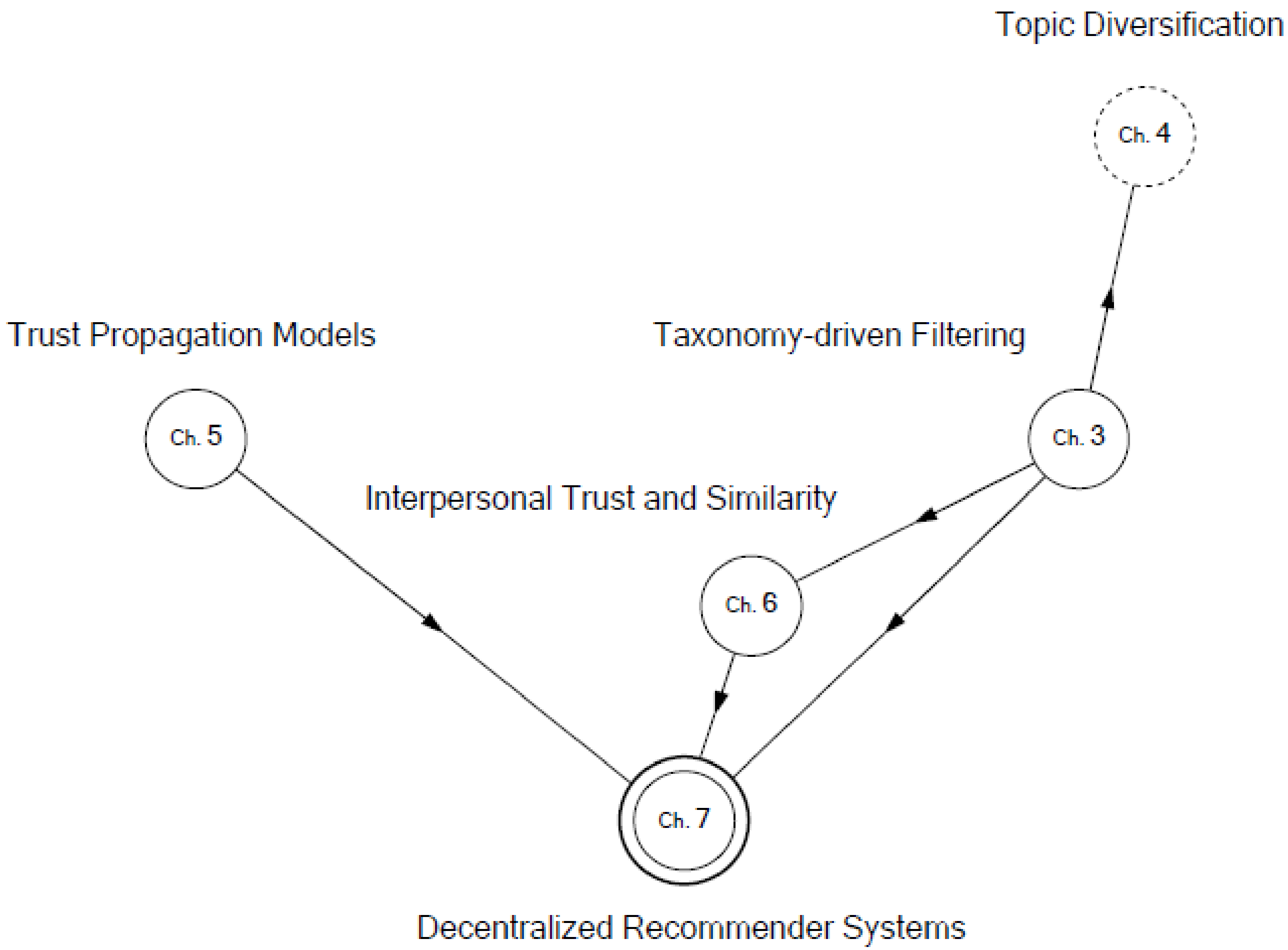
*Spreading Activation Models for Trust Propagation*

- 2003~2005: PhD student at Albert-Ludwigs-University Freiburg, Germany
- 2005~2007: Consultant with Siemens AG
- 2008~2010: Consultant with BCG
  while postdoc with Albert-Ludwigs-University Freiburg

# Overview of Ziegler's work

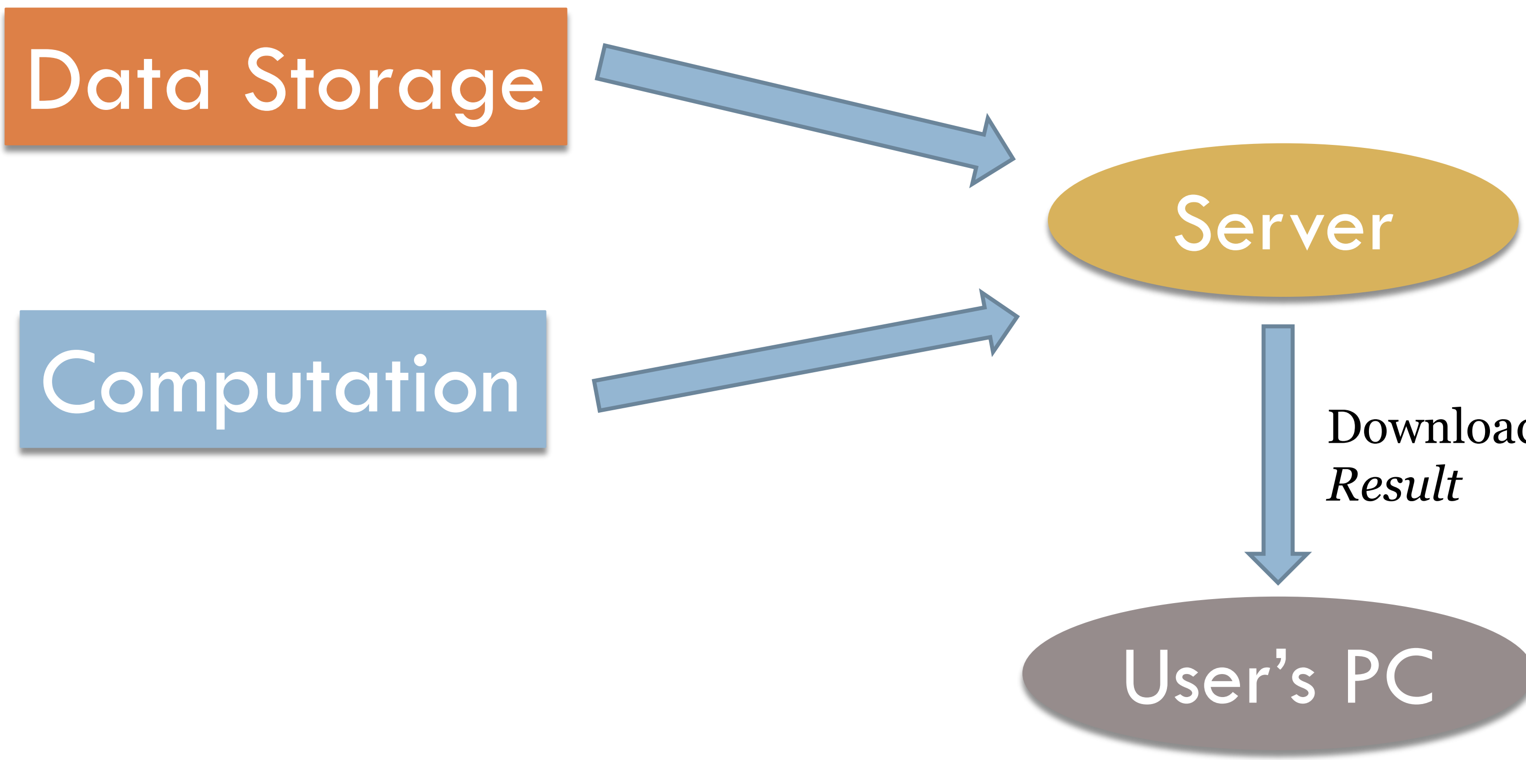
- Taxonomy-driven filtering
  - Topic diversification
- Trust propagation model

# Overview of Ziegler's work



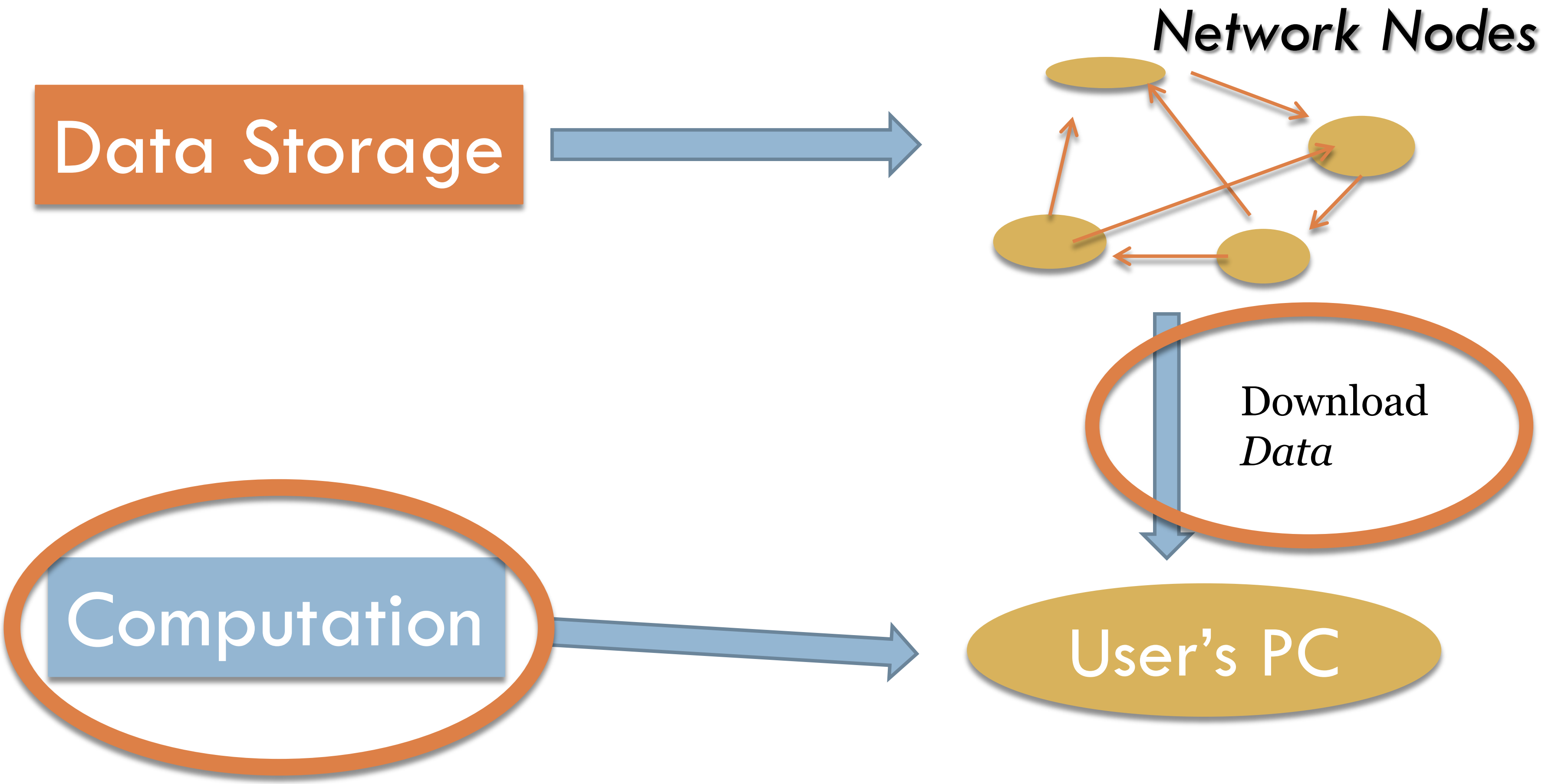
# About *Decentralization*

- Centralized *(traditional)*



# About *Decentralization*

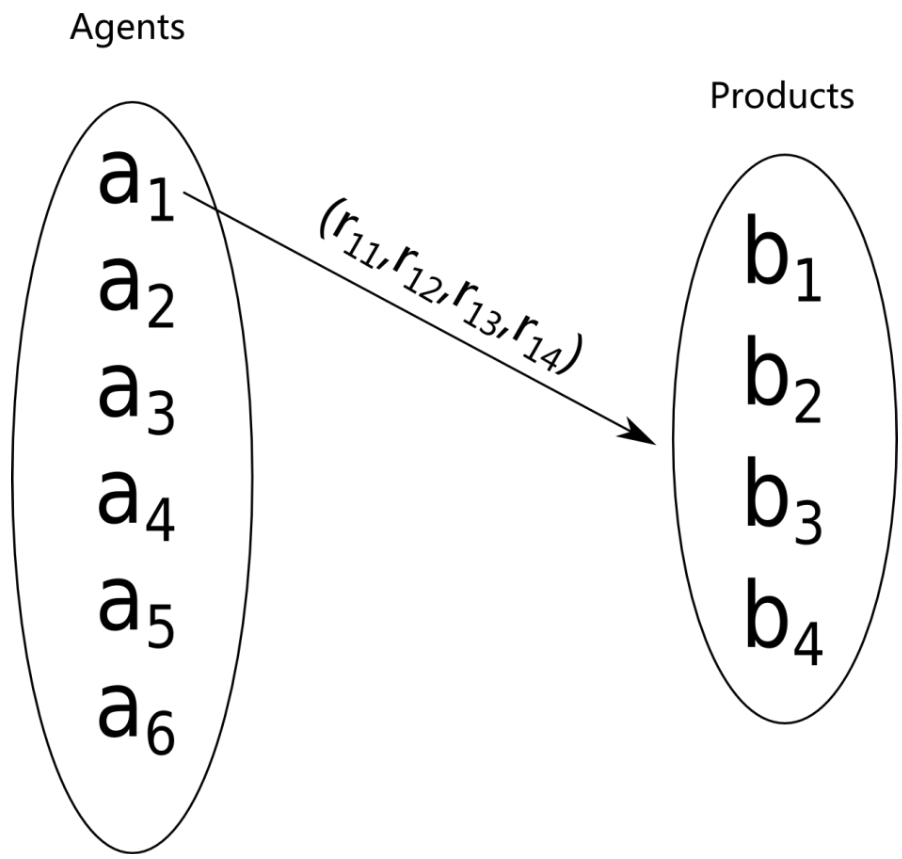
- Decentralized



# Taxonomy-driven Filtering

- Intuition

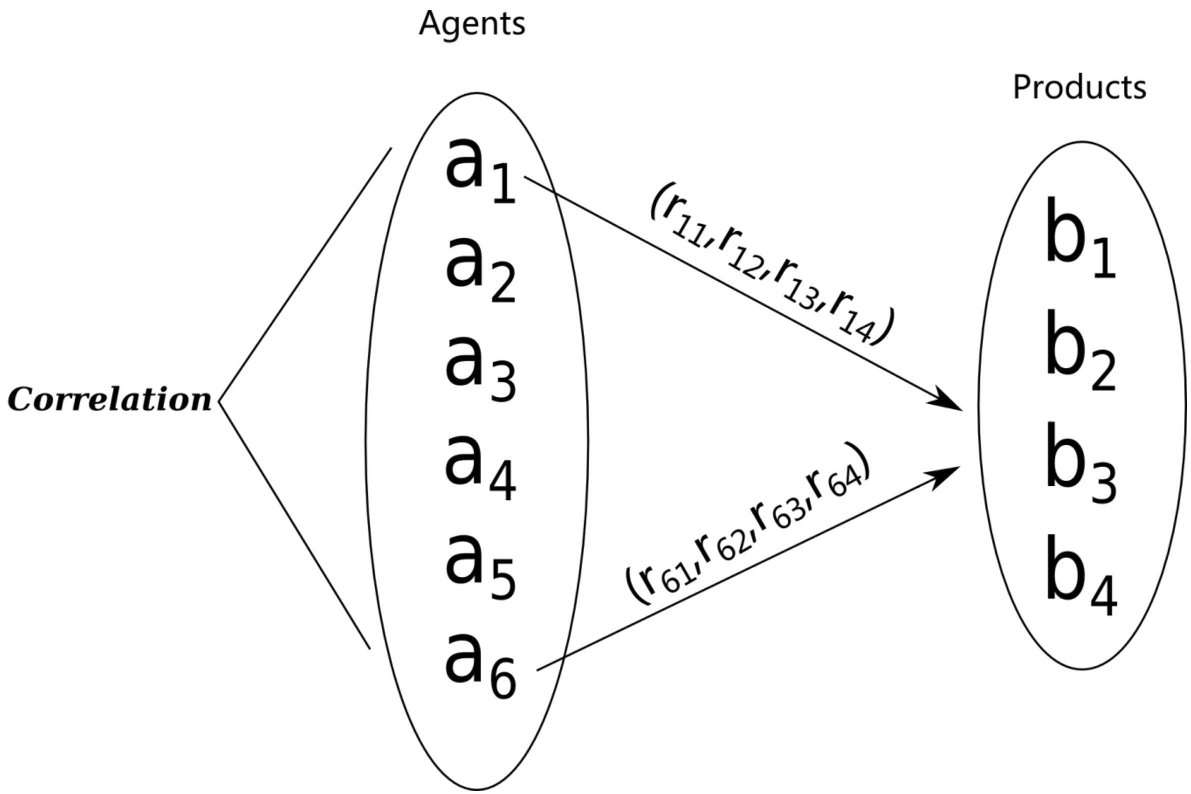
1. *Profiling*



# Taxonomy-driven Filtering

- Intuition

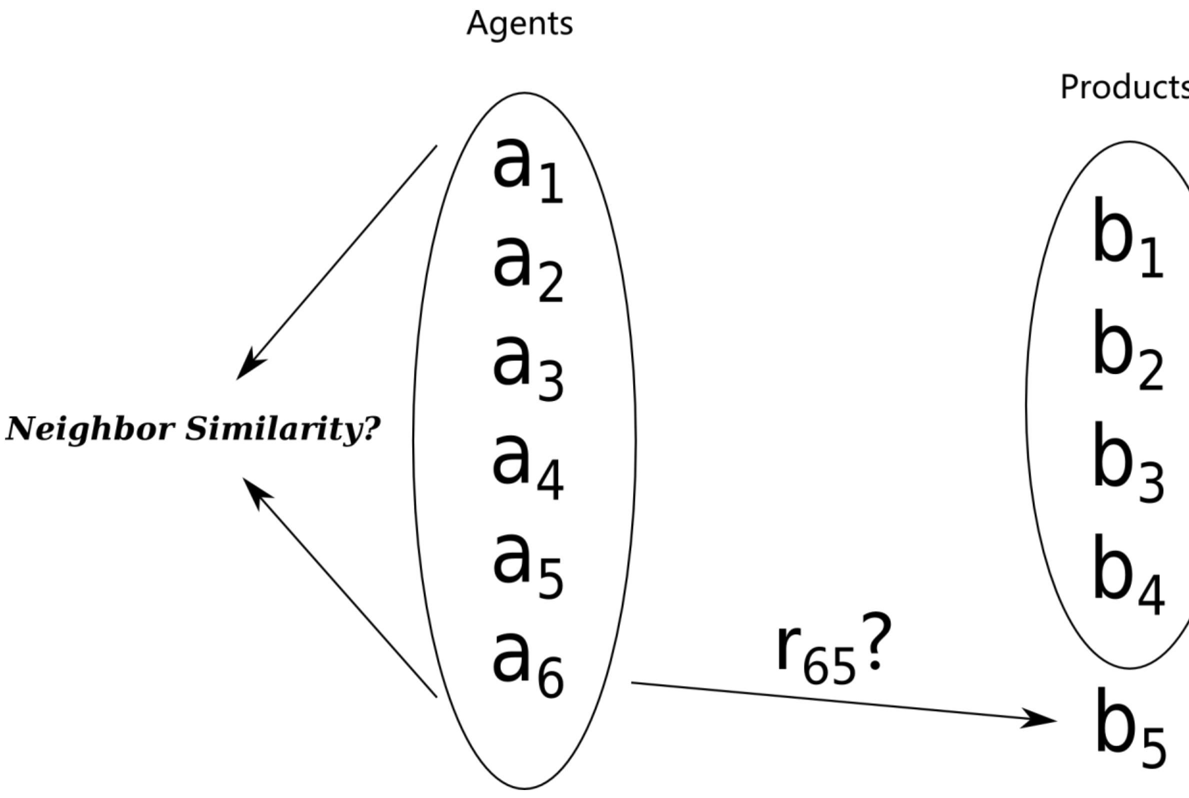
2. *Proximity computation and Neighborhood formation*



# Taxonomy-driven Filtering

- Intuition

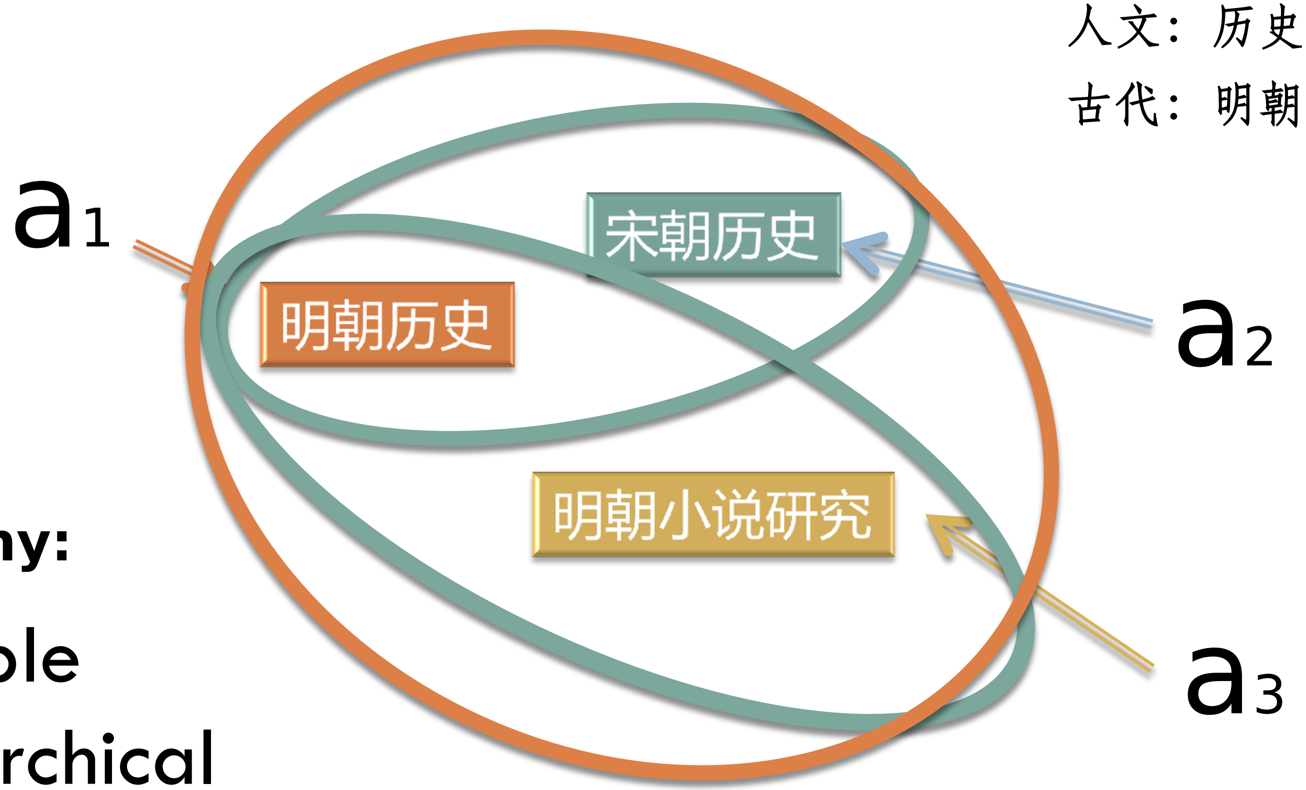
3. *Rating prediction and Recommendation generation*



# Taxonomy-driven Filtering

- Problem — Sparsity (low profile overlapping)

人文：历史：明朝历史
古代：明朝



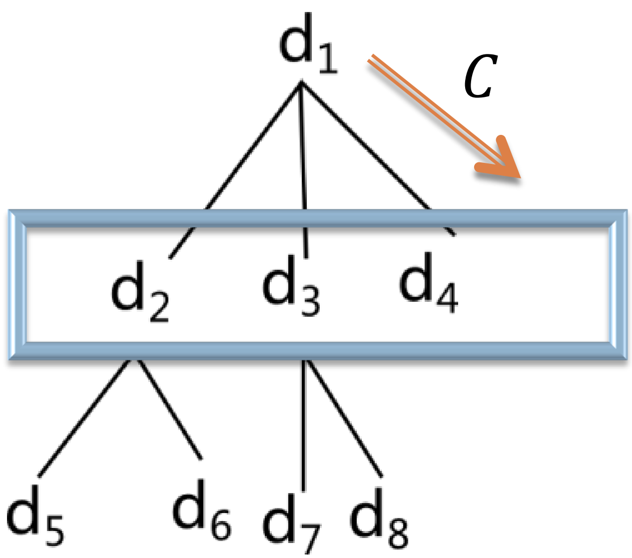

**Taxonomy:**

- Multiple
- Hierarchical

# Taxonomy-driven Filtering

- Mathematical Model
  - Tree-structural Taxonomy Set

$$D = \{d_1, d_2, d_3, \cdots, d_l\}$$



# Taxonomy-driven Filtering

- Mathematical Model
  - Taxonomy-based (topic-based) profile

$$\vec{v}_i = \left(v_{i,1}, v_{i,2}, \cdots, v_{i,l}\right)$$

  - Normalization

(we only care user's *interest distribution among all topics*)

$$\sum_{k=1}^{l} v_{i,k} = s$$

# Taxonomy-driven Filtering

- Mathematical Model
  - Generating profile from *implicit ratings*
    - Assumption: implicit ratings are expressed in *binary form* *(discuss it later)*

For each user $a_i$

$$B = \{b_1, b_2, b_3, \cdots, b_m\}$$

$$R_i \subseteq B$$

*Rated(positively)*

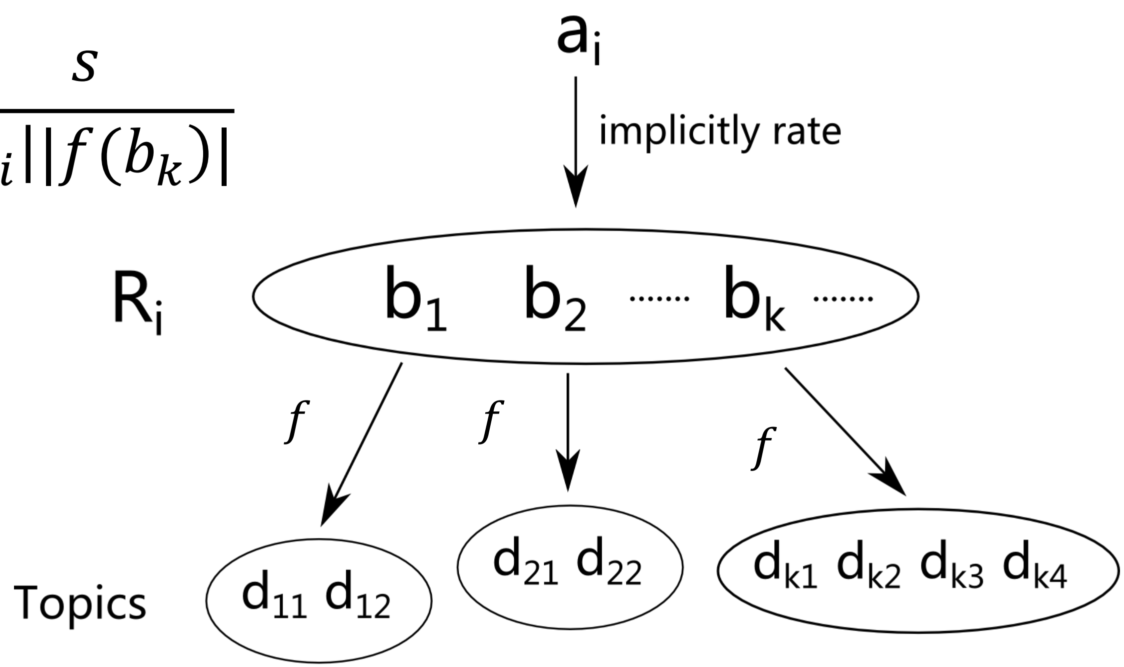
$$B / R_i$$

*Unrated*

# Taxonomy-driven Filtering

- Mathematical Model
  - Generating profile from *implicit ratings*
    - Step1: Distribute total score evenly among topics that rated products belong to

$$sc(d_{ke}) = \frac{s}{|R_i||f(b_k)|}$$



# Taxonomy-driven Filtering

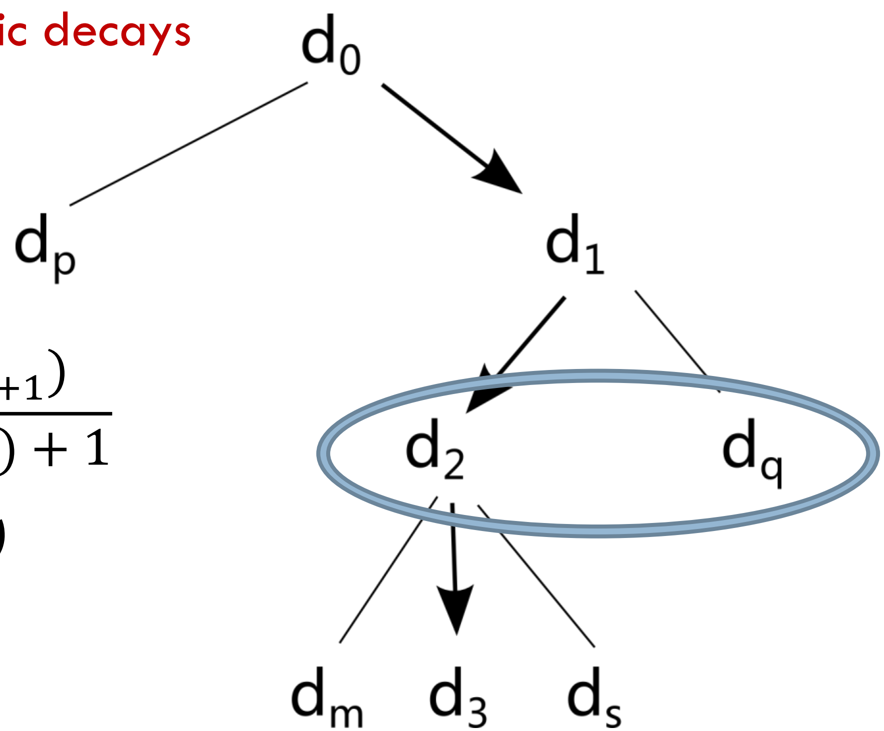
- Mathematical Model
  - Generating profile from *implicit ratings*
    - Step2: Redistribute the score to all topics with two assumptions
      - Conservation along hierarchical path (to ensure normalization)

$$\sum_{m=0}^{q} sco(p_m) = sc(d_{k_e})$$

      - Redistribute each topic score along hierarchical path to its super-topics with semantic decays

# Taxonomy-driven Filtering

- Generating profile from *implicit ratings*
  - Step2: Redistribute the score to all topics with two assumptions
    - Conservation(normalization)
    - Redistribute each topic score along hierarchical path to its super-topics with semantic decays




$$sco(d_m) = k\,\frac{sco(d_{m+1})}{sib(d_{m+1}) + 1}$$

***(Recursive redistribution)***

# Taxonomy-driven Filtering

- Mathematical Model
  - Generating profile from *implicit ratings*
  - Measuring proximity and forming neighborhood
    - Proximity measurement: *Pearson Correlation*

$$c(a_i, a_j) = \frac{\sum_{k=1}^{|D|} (\vec{v}_{i,k} - \bar{v}_i)(\vec{v}_{j,k} - \bar{v}_j)}{\sqrt{\sum_{k=1}^{|D|} (\vec{v}_{i,k} - \bar{v}_i)^2} \sqrt{\sum_{k=1}^{|D|} (\vec{v}_{j,k} - \bar{v}_j)^2}}$$

    - Neighbors Selection: *Top-M*

# Taxonomy-driven Filtering

- Mathematical Model
  - Generating profile from *implicit ratings*
  - Measuring proximity and forming neighborhood
  - Forming recommendation list
    - Two-fold relevance (user-proximity & product-proximity)
      - User-proximity: Pearson
      - Product-proximity: *"dummy user trick"*

*dummy user* $a_\theta \; with \; R_\theta = \{b_k\}$

$$c_b(a_i, b_k) \triangleq c(a_i, a_\theta)$$

# Taxonomy-driven Filtering

- Mathematical Model
  - Generating profile from *implicit ratings*
  - Measuring proximity and forming neighborhood
  - Forming recommendation list
    - Two-fold relevance (user-proximity & product-proximity)

$$w_i(b_k) = \frac{qc_b(a_i, b_k) \cdot \sum_{a_j \in A_i(b_k)} c(a_i, a_j)}{|A_i(b_k)| + \Upsilon_R}$$

***For fine-tune***

# Taxonomy-driven Filtering

- Mathematical Model
  - Generating profile from *implicit ratings*
  - Measuring proximity and forming neighborhood
  - Forming recommendation list
    - Two-fold relevance (user-proximity & product-proximity)
    - *Topic Diversification Technique*
      - For alleviating *"Winners Take All"* problem

# Taxonomy-driven Filtering

- Topic Diversification

$$OriginalRank(b) + \Theta_F DissimilarRank(b)$$

***Sort***

*Recommendation List*

# Taxonomy-driven Filtering

- Mathematical Model
- Evaluation
  - How to evaluate the utility of a recommender
    - Precision
    - Recall
    - *Breese Score*

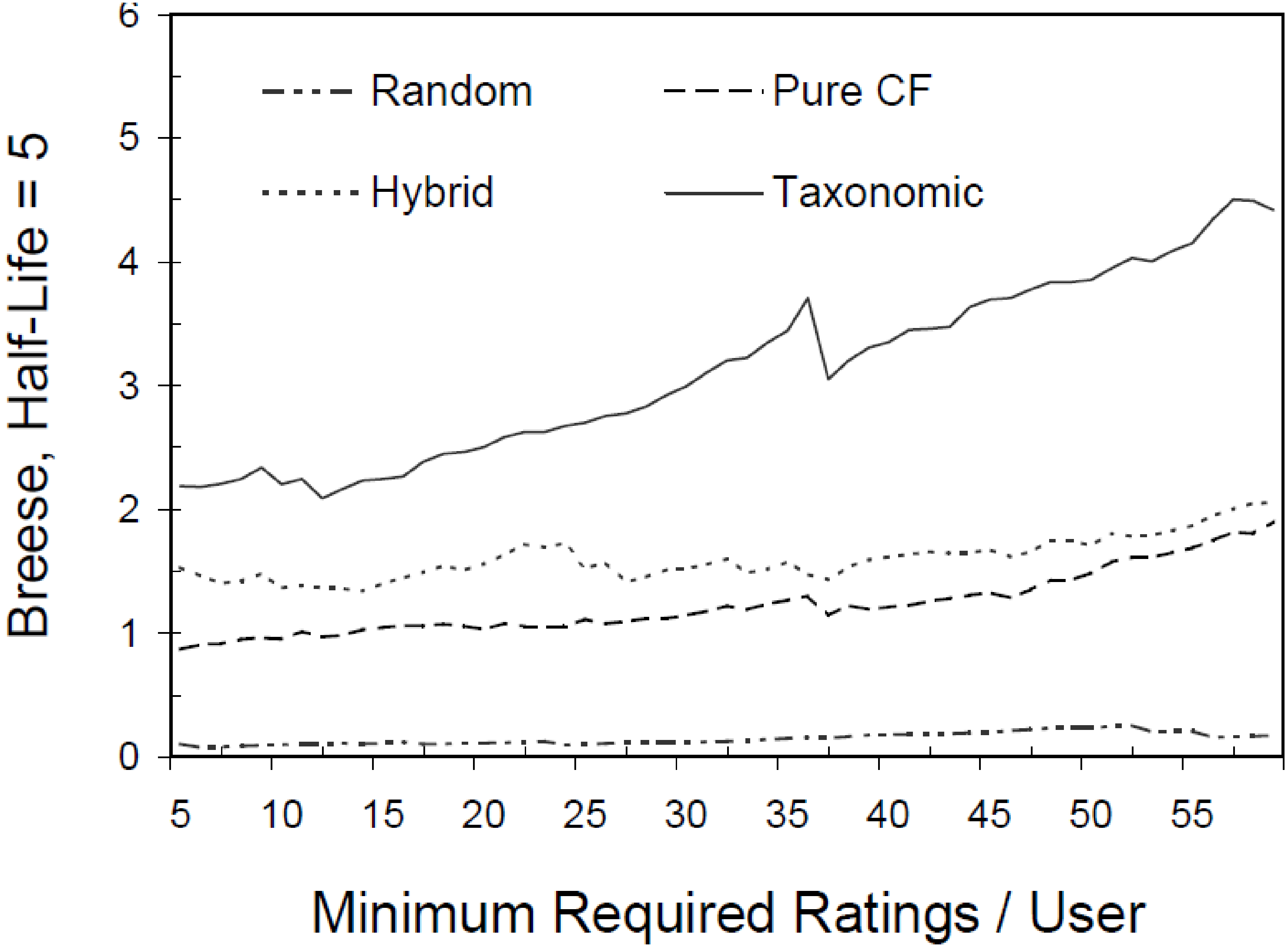
    *the expected utility of a recommendation list is simply the probability of viewing a recommended product that is actually relevant*

    *Assumption: each successive item in the recommendation list is less likely to be viewed with exponential decay*

# Taxonomy-driven Filtering

- Mathematical Model
- Evaluation
  - How to evaluate the utility of a recommender
  - Comparison: with random recommender and two other CF's
  - Two sets of empirical data
    - *All Consuming book community (sparse)*
    - *MovieLens movie community (dense)*
  - Result
    - Work better than counterparts in comparison
    - *Larger advantage* gap in *sparser* community

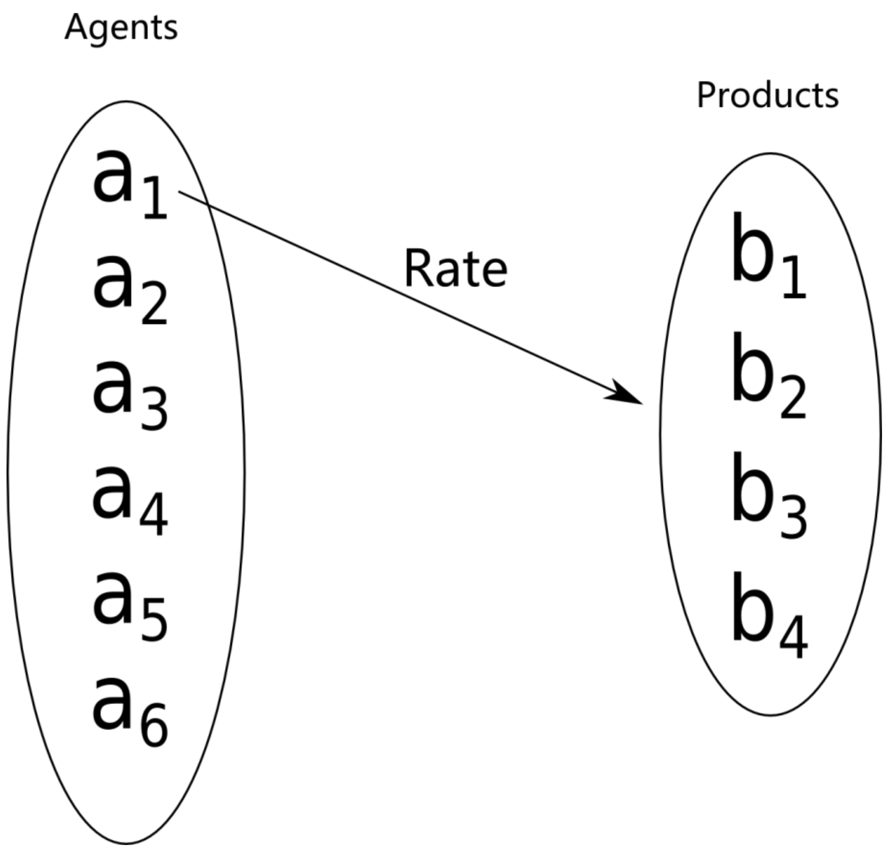
# Taxonomy-driven Filtering



# My Review on Taxonomy-driven Filtering

- Effectiveness of Taxonomy
  - *Structuralize user-rating network*

Form neighborhoods




Categorize with taxonomy

# My Review on Taxonomy-driven Filtering

- Recommendation *beyond the dimension of accuracy*?
  - *Topic Diversification*
    - *Positive-feedback of recommendation list? (initial-value sensitive?)*
  - *How would recommendation be adaptive in response to user's behavior/attitude?*
    - *What can I do if I'm not satisfied with recommendation?*

豆瓣FM
douban.fm

# My Review on Taxonomy-driven Filtering

- *Explicit* rating vs. *Implicit* rating

(*Mr. Ziegler made his recommender based on implicit*)

  - Is explicit rating *necessarily a burden* upon users? 

  - Can implicit rating really *supersede* explicit rating?
  - Is implicit rating necessarily *binary*?

$$a_i \rightarrow R_i = \{rated\ products\}$$

    - Observation *diversity*
      - Viewing/purchase, browsing behavior, listened/watched, comment etc.
    - *Insufficient input information?*

$$unrated \neq dislike$$

# My Review on Taxonomy-driven Filtering

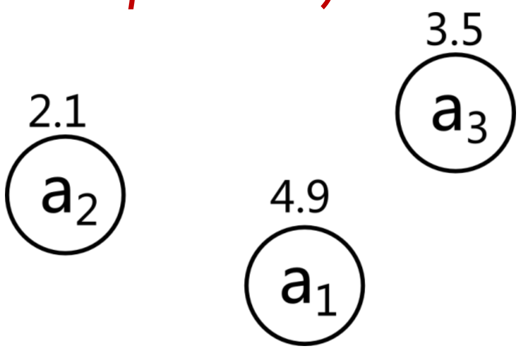
- What do we *rely on* when *designing* a recommender?
  - *What knowledge do we have about users' behavior?*
  - *Are statistical distributions useful?*

- What do we *rely on* when *evaluating* a recommender?
  - Can *one set of data* justify the performance for *all data*?
  - *Is simulation completely unreliable/impossible?*

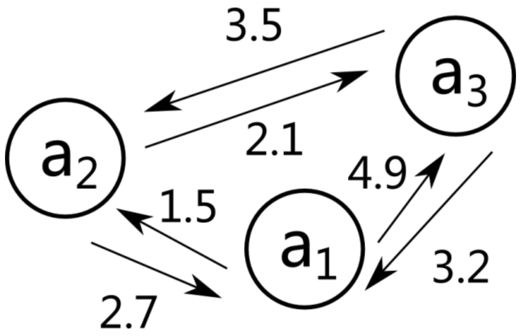
- *How to fine-tune parameters?*

# Trust Propagation Model

- Computational trust in three dimensions
  - Network perspective
    - *Global trust* *(public reputation)*

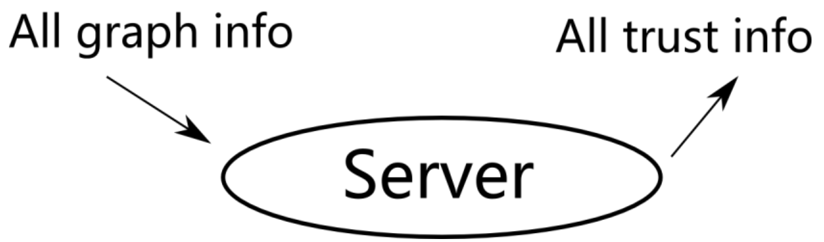



    - *Local trust* *(one's personal trust towards another)*

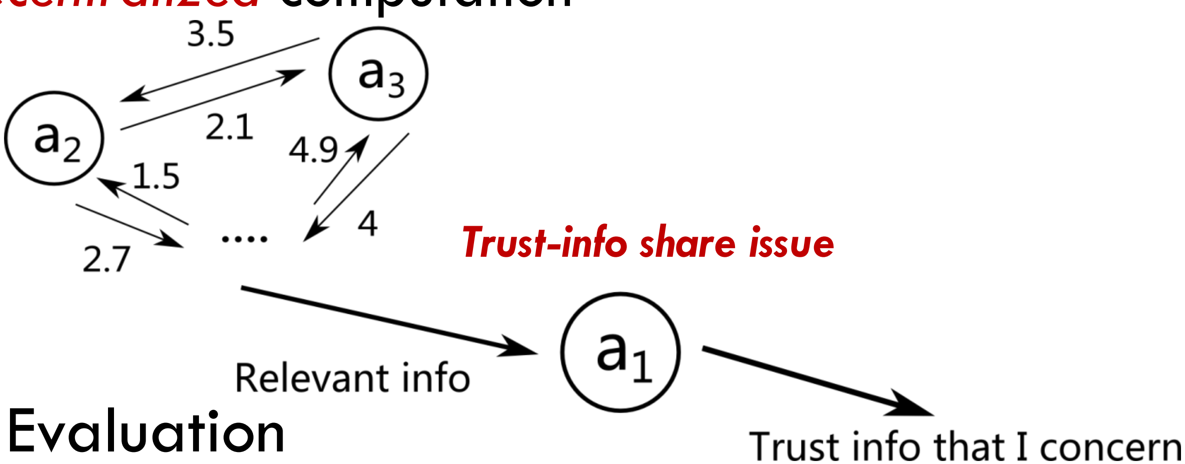


# Trust Propagation Model

- Computational trust in three dimensions
  - Computation Locus
    - *Centralized computation*




    - *Decentralized* computation




  - Link Evaluation

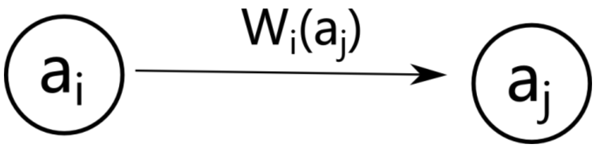
# Trust Propagation Model

- Appleseed trust model (*energy-propagation*)
  - Trust assertions

$$A = \{a_1, a_2, \cdots, a_n\}$$

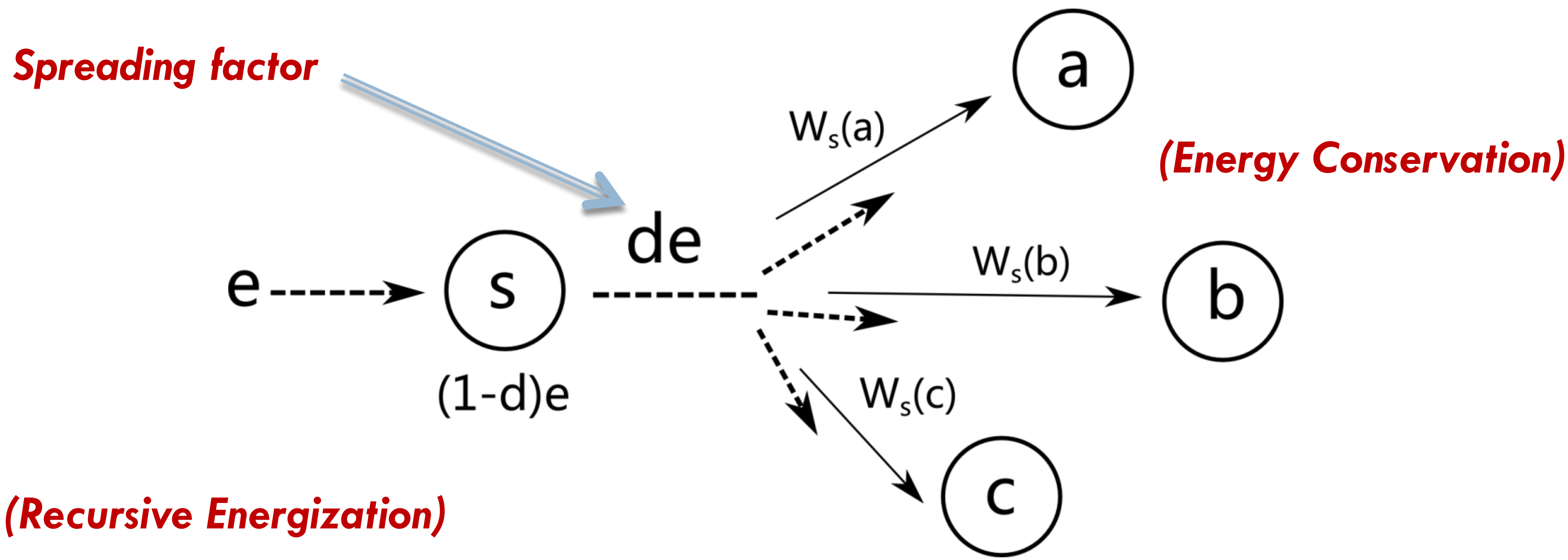
Agent i's trust assertions

$$W_i : A \rightarrow [0,1]^{\perp}$$



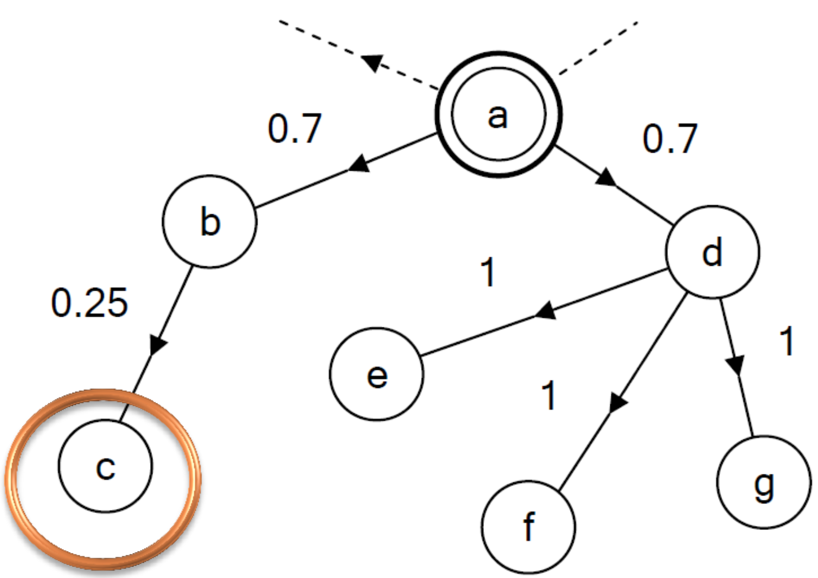
# Trust Propagation Model

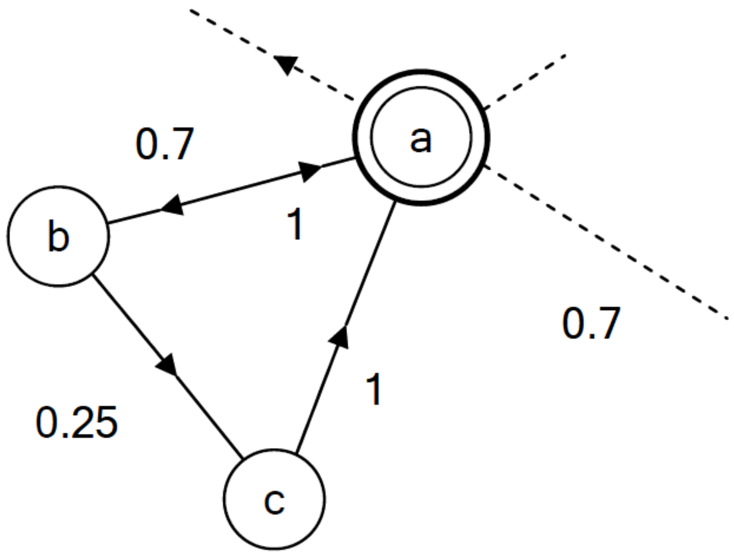
- Appleseed trust model (*energy-propagation*)
  - Trust assertions
  - Trust propagation (*energize*)



# Trust Propagation Model

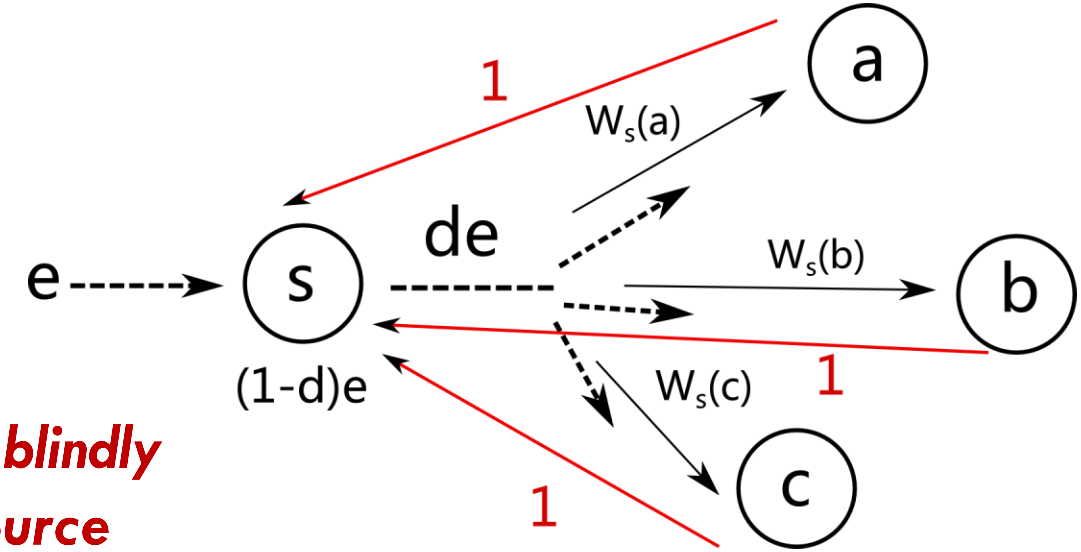
- Appleseed trust model (*energy-propagation*)
  - Trust assertions
  - Trust propagation (*energize*)
    - Problems



# Trust Propagation Model

- Appleseed trust model (*energy-propagation*)
  - Trust assertions
  - Trust propagation (*energize*)
    - Problems
    - Solution — ***Backward trust propagation to trust source***




***Every node blindly trusts the source***

# Trust Propagation Model

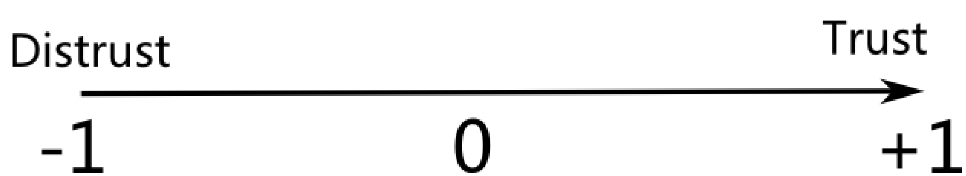
- Appleseed trust model (*energy-propagation*)
  - Trust assertions
  - Trust propagation (*energize*)
  - Computation
    - Only *relevant data* are acquired for local computation
    - Terminate recursion when reaching threshold
      - ***Convergence***

$$\sum_{x \in A} trust_i(x) \uparrow \leq InitialEnergy$$

# Trust Propagation Model

- Appleseed trust model (*energy-propagation*)
  - Trust assertions
  - Trust propagation (*energize*)
  - Computation
  - Analysis
    - Attack-resistance

      *"Bad node's trust assignment cannot affect trust value."*
    - Incorporate ***distrust*** into this model?



# Trust Propagation Model

- Appleseed trust model (*energy-propagation*)
  - Incorporate ***distrust*** into this model?

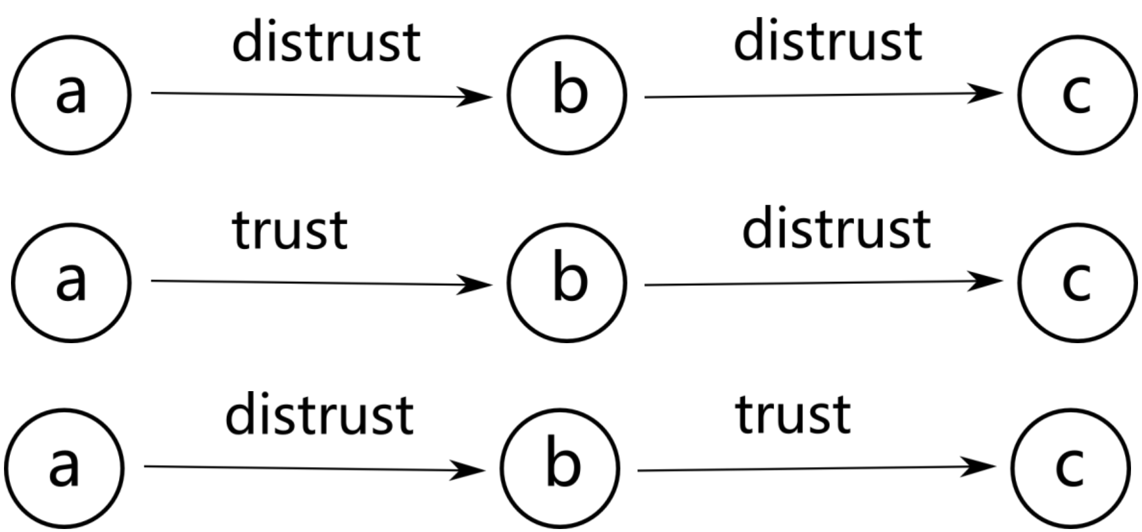
$$e_{x\to y} = \begin{cases} d \cdot in(x) \cdot \dfrac{|W(x,y)|}{\sum_{(x,s)\in E} |W(x,s)|}, \; in(x) > 0 \\ 0, \; in(x) \leq 0 \end{cases}$$

# My Review on Trust Propagation Model

- Further investigation into *trust/distrust interaction* **?**
  - What does trust/distrust mean?

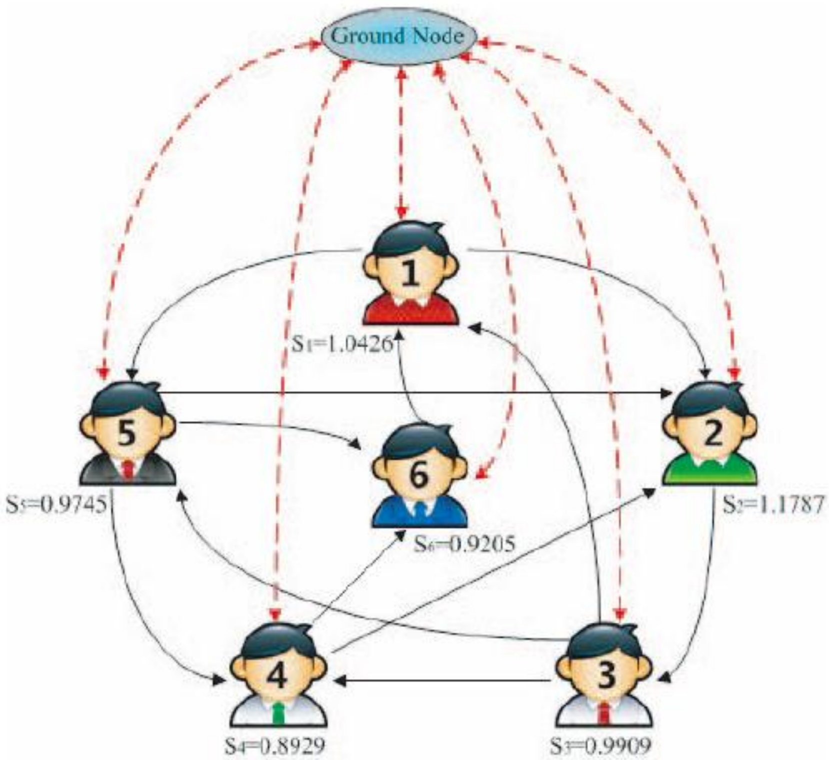
$$distrust \neq lack\ of\ trust\ \text{info}$$

  - How do trust and distrust *propagate and interact ?*



# My Review on Trust Propagation Model

- The implication of introducing a super-node *connecting all other nodes ?*
  - *also applied in "Leaders in Social Networks"*



# Thanks!

*Questions and Discussion …*

by Richard Kwo
richardkwo@gmail.com