The Lifecycle and Cascade of WeChat Social Messaging Groups

Jiezhong Qiu† , Yixuan Li#, Jie Tang† , Zheng Lu‡ , Hao Ye‡ , Bo Chen‡,
Qiang Yang* and John E. Hopcroft#
† Tsinghua University
# Cornell University
‡ Tencent Corporation, Beijing, China
* Hong Kong University of Science and Technology

Social Media

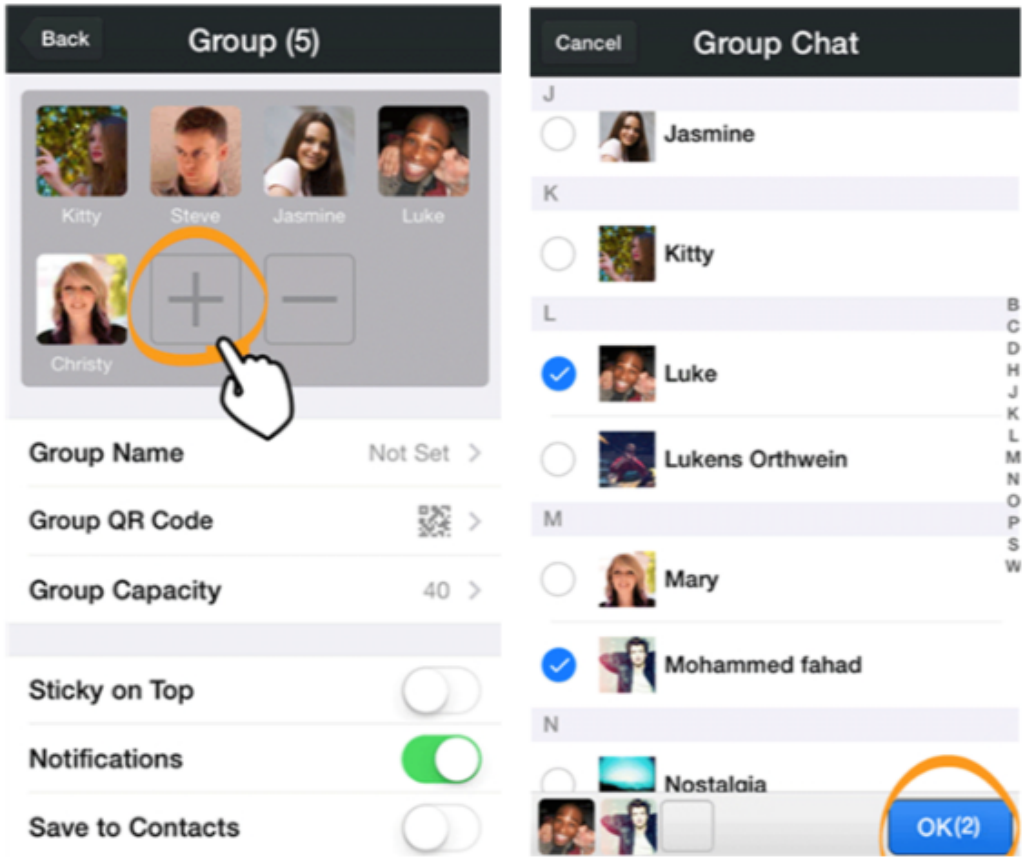
• Open, Fast, Visible

• Private, Relationship-focused

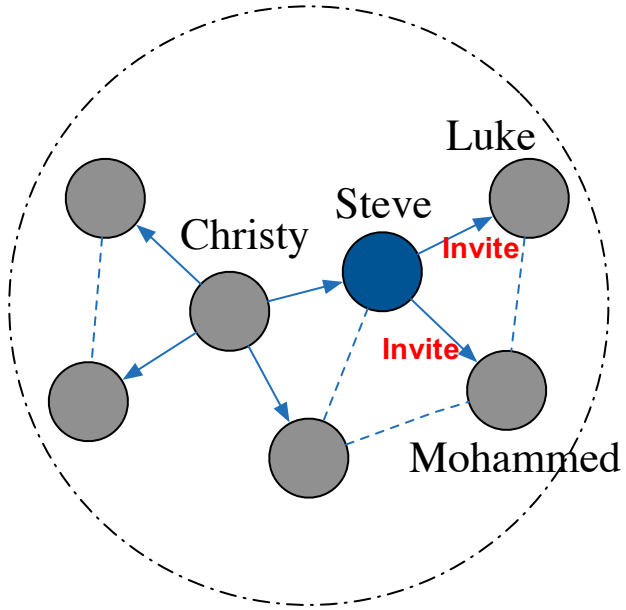
•>1 billion created accounts
•~697 million MAUs
• >70 million MAUs outside of China

http://tencent.com/en-us/ir/news/2016.shtml

Group Chat in WeChat



(a) WeChat group membership (b) Membership invitation



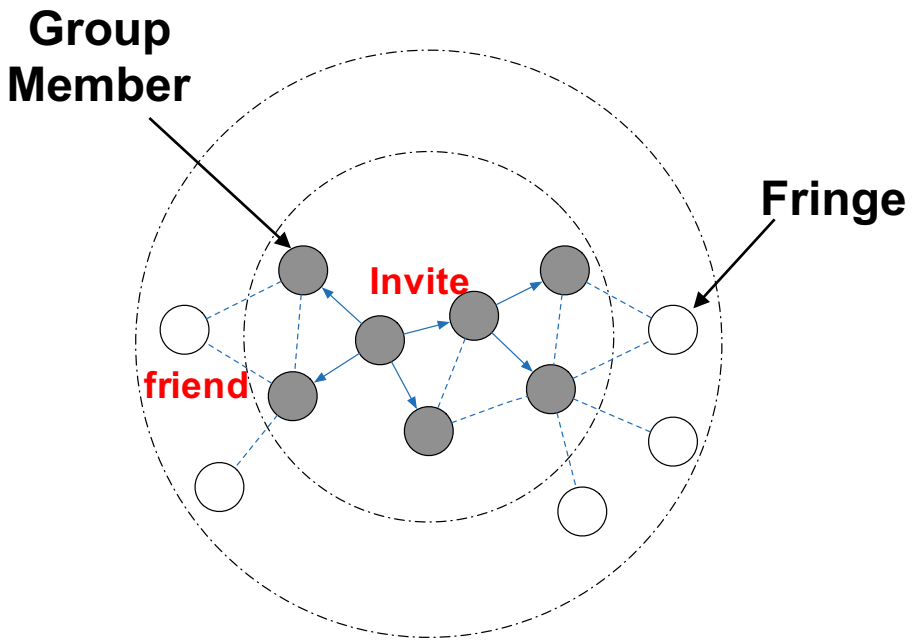
- **~2.3 million groups generated everyday**
- **>25% messages are generated in group chats**

Data Set

Table 1: Summary of data set.

Category	Type	Number
Group	Total	474,726
	Min group size	3
	Max group size	500
User	Total	245,352,140
Invitation	Total	2,013,351
Friendship	Total	624,529,005

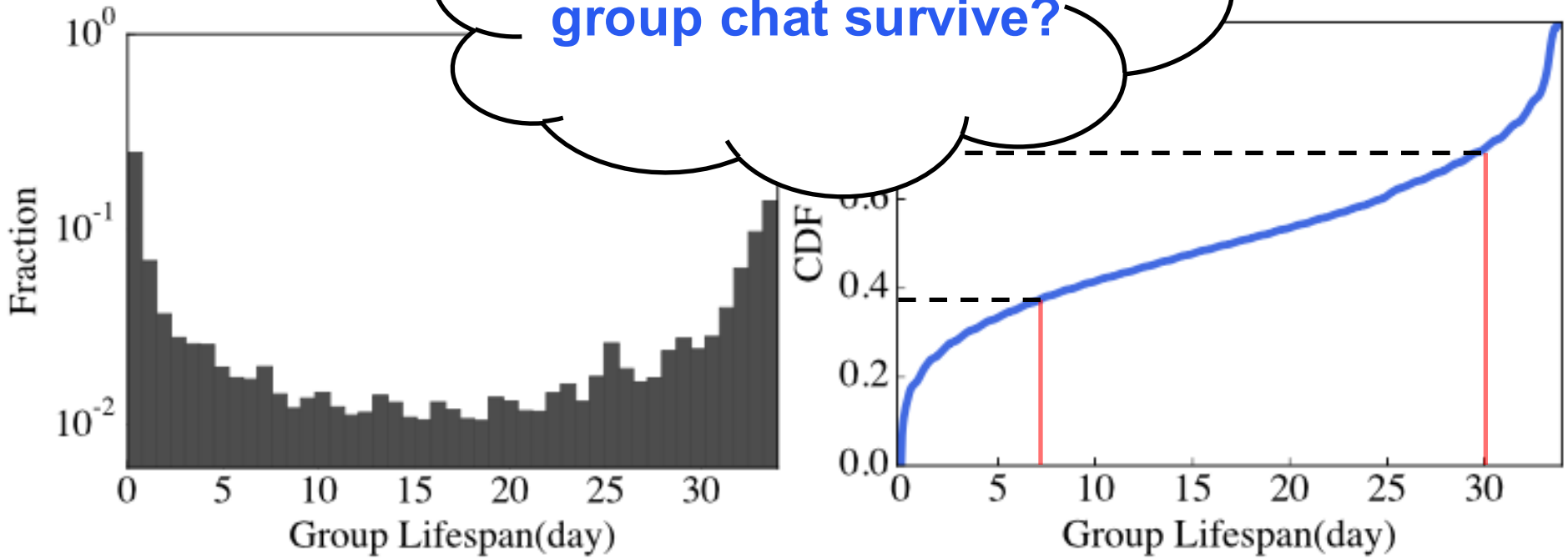
- **Group: groups generated on July 26th, 2015**
- **User: group members + users in fringe**
- **Invitation: (u, v, C, T)**
- **Friendship: (u, v, T)**



Group Lifecycle Dichotomy

Definition: **Group Lifespan. Duration from the timestamp at which a group is initialized, to th… …oup members sends chat messag…**

How long would a group chat survive?



Short-term group v.s Long-term group

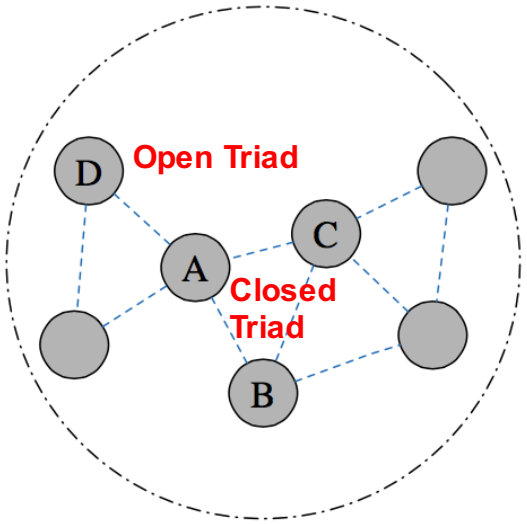
Group Lifecycle Dichotomy – Case Study

Table 2: Case study by group displayed name.

Category	Long	Short	Example
Travel	0	8	Discuss on a short trip
Meeting	1	2	Schedule an official meeting
Event	4	13	Plan a wedding
Entertain	5	13	Dine together
Organization	9	0	Departments of company
Class	12	4	Course for GRE test
Friend	13	0	Childhood friend
Family	16	0	A family of three

Short-term group v.s Long-term group
Event-driven v.s. Relationship-driven

Group Lifecycle Dichotomy – Structure Dynamics

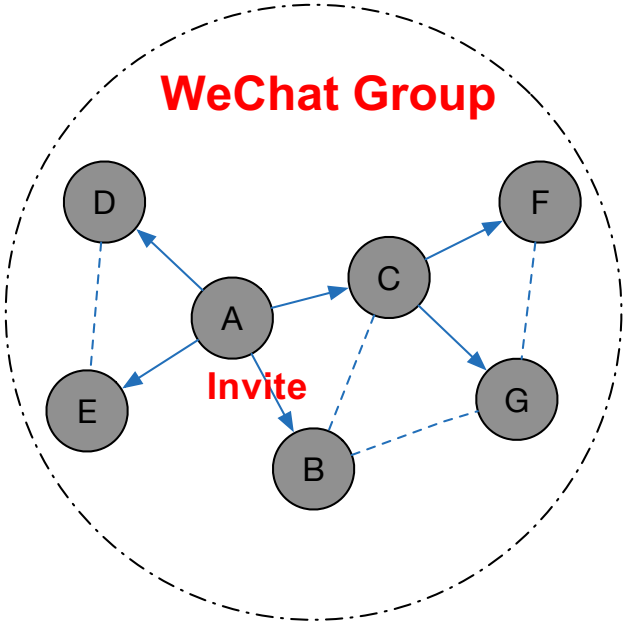


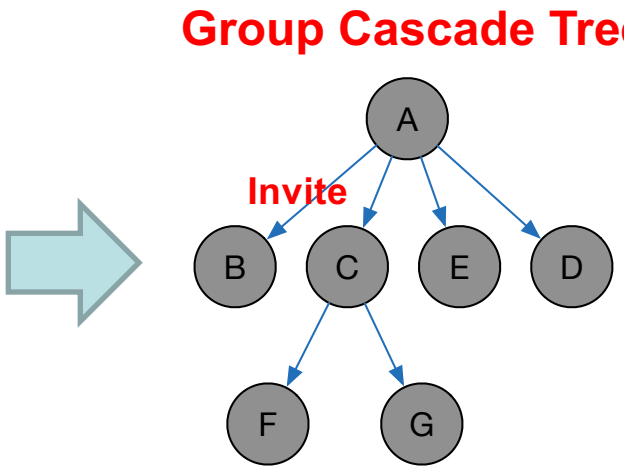
(a) Example

- **Long-term Group: Strong dynamics in terms of underlying friendship structure.**
- **Short-term Group: Less likely to develop friendship over time.**

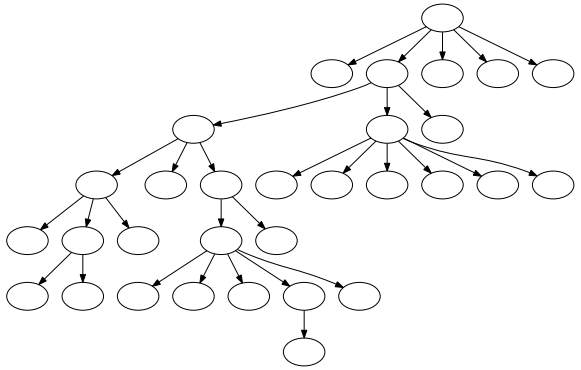
Group Lifecycle Dichotomy – Group Cascade Tree

Definition: **Group Cascade Tree. A directed graph where each group member is a node, and a directed edge from u to v is constructed if u (inviter) successfully invites v (invitee) to the group.**

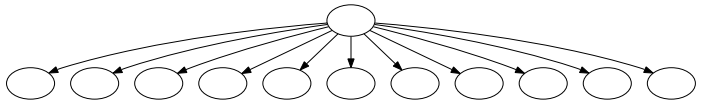




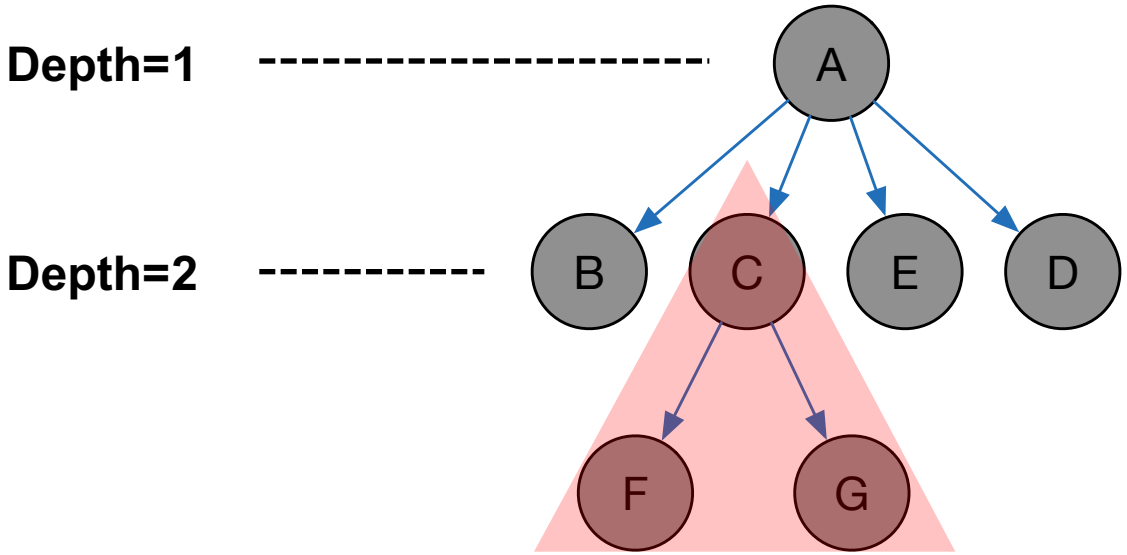
Example of long-term groups



Example of short-term groups



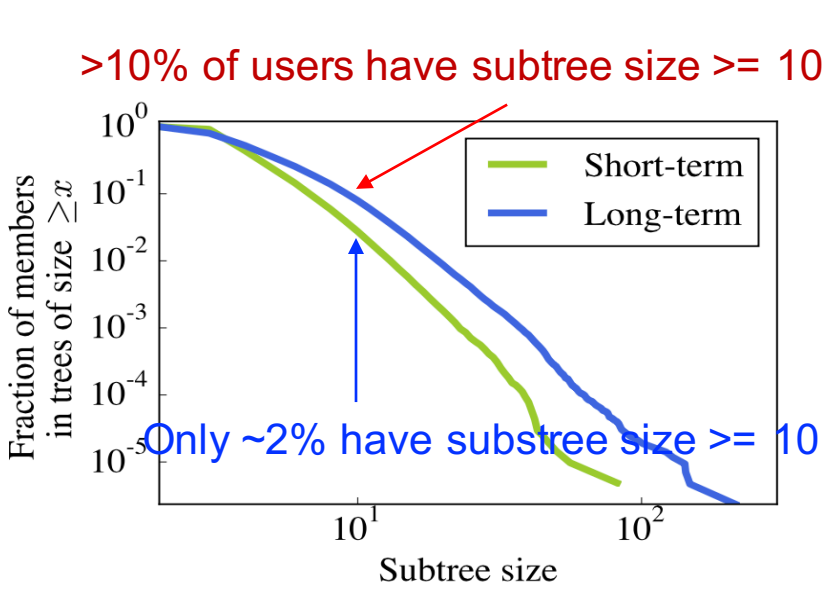
Group Lifecycle Dichotomy – Cascade Tree Pattern



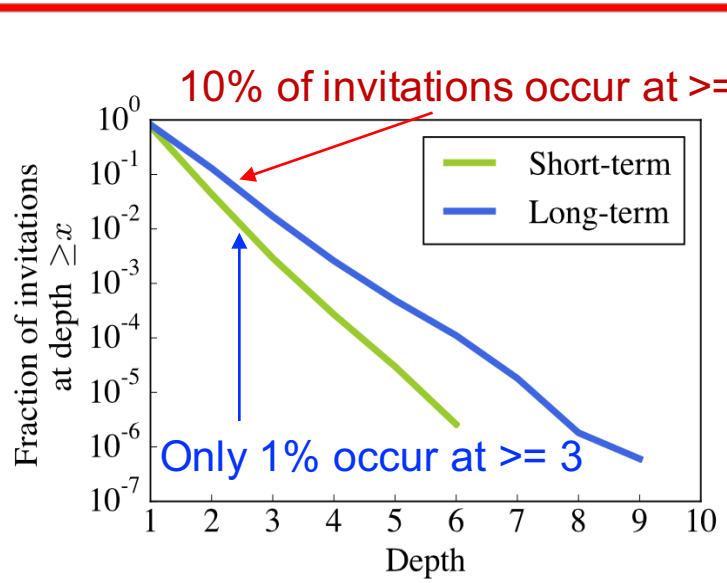
$$\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}$$

- **Subtree size: The size of sub-cascade**
- **Depth: The depth of invitation**
- **Wiener Index: Average distance between two nodes**

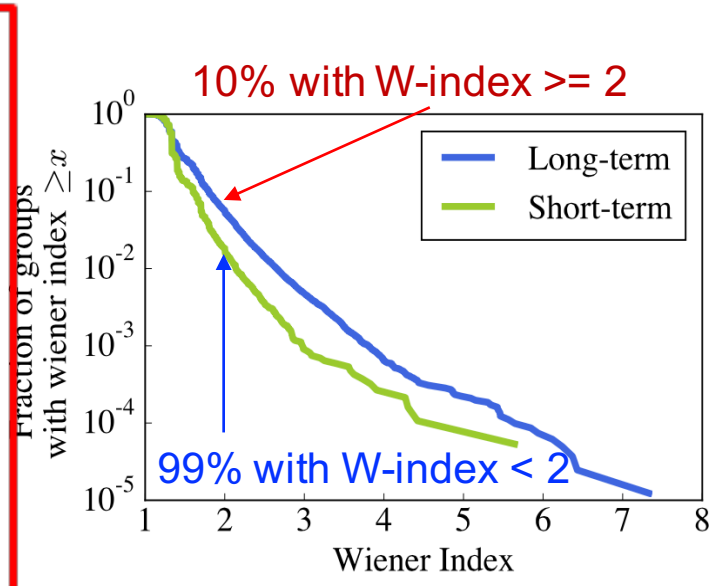
Group Lifecycle Dichotomy—Cascade Tree Pattern



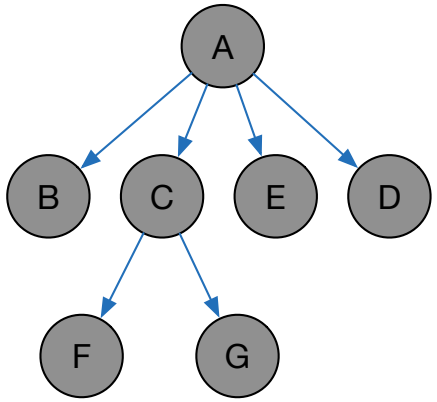
(b) Subtree size



(c) Depth



(d) Wiener index



- **For node C**
  - **Subtree size: 3**
  - **Depth: 2**
- **For the left example:**
  - **Wiener index: 2**

Group Lifecycle Dichotomy — Features

Group Level:
For group C at time T

Group Structure	The number of open triads at T and at the setting up of group.
	The number of closed triads at T and at the setting up of group.
Cascade Tree	Wiener index.
	Number of members whose depth equal to k, k = 1,2,...,9.
Demographics	Number of members who stated their gender to be X.
	Entropy of member's gender

Group Lifecycle Dichotomy—Prediction

SVM 10-fold Cross Validation

Features	AUC	Precision	Recall	F1
All Features	**66.62**	**63.23**	**57.66**	**60.32**
-Structure	64.75	59.36	62.83	61.04
-Cascade	65.36	64.49	47.67	54.82
-Demographics	65.24	57.35	65.71	61.25
+Structure	64.21	61.98	42.51	50.43
+Cascade	**61.23**	**57.35**	**65.71**	**61.25**
+Demographics	62.77	63.18	41.41	50.03

- **Task 1: Group Separability: Predict groups' lifespan.**
- **Task 2: Early Prediction: Can we predict the group lifecycle in early stage.**

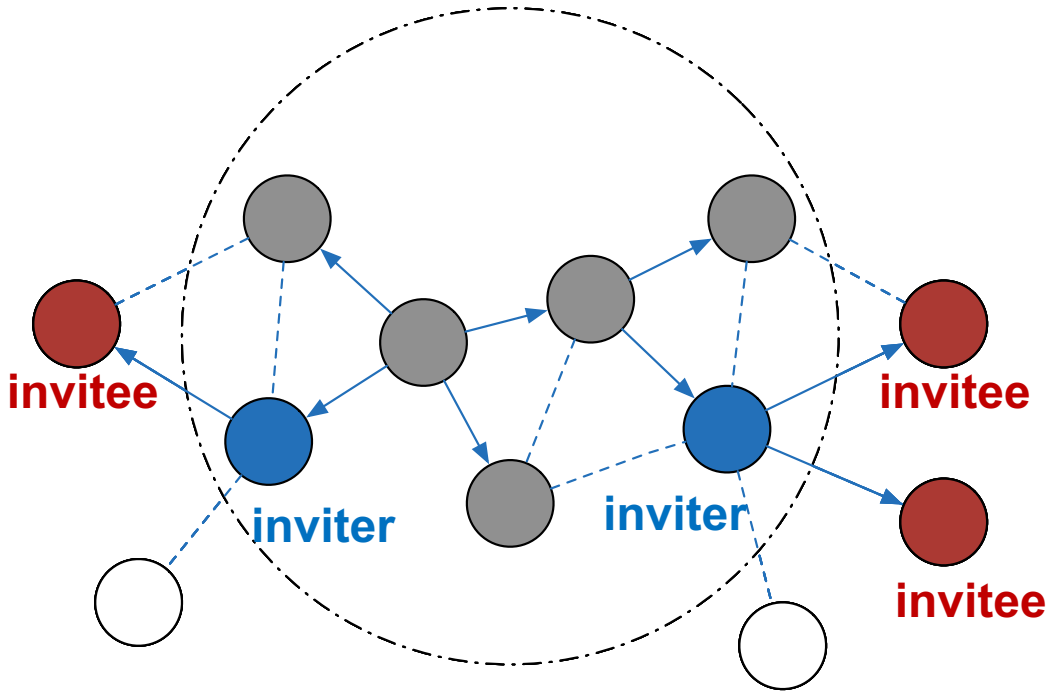
Group Lifecycle Dichotomy—Prediction

SVM 10-fold Cross Validation

Features	AUC	Precision	Recall	F1
1 hour	57.95	54.16	56.80	55.45
1 day	**65.08**	**61.92**	**53.38**	**57.34**
5 days	65.46	62.52	54.11	58.01
10 days	65.57	62.48	56.81	59.51
20 days	65.76	62.78	56.56	59.51
1 month	**66.62**	**63.23**	**57.66**	**60.32**

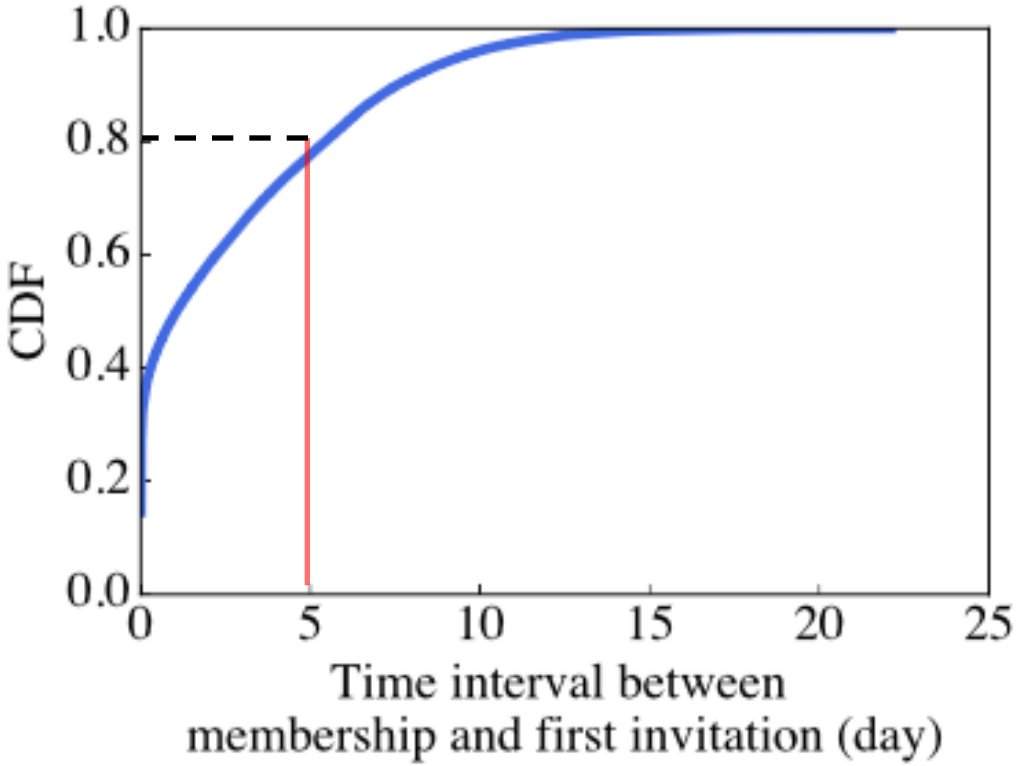
- **Task 1: Group Separability: Predict groups' lifespan.**
- **Task 2: Early Prediction: Can we predict the group lifespan in early stage.**

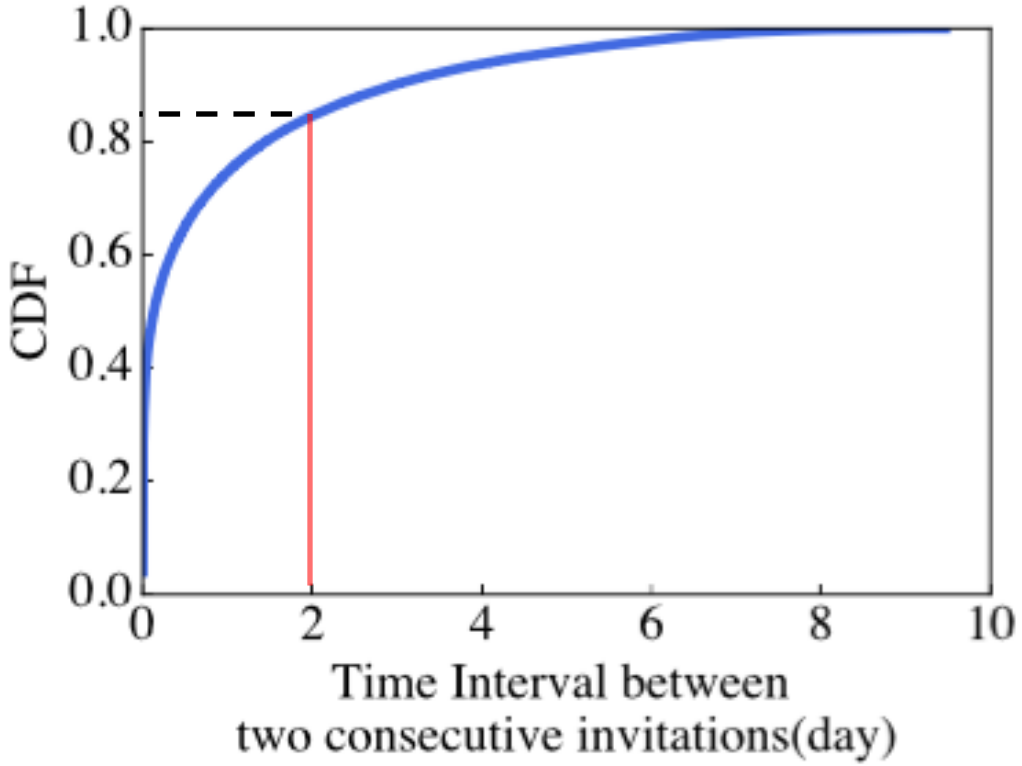
Membership Cascade



- Q1: Who are inviters?
- Q2: Who are invitees?

Membership Cascade—Inviter

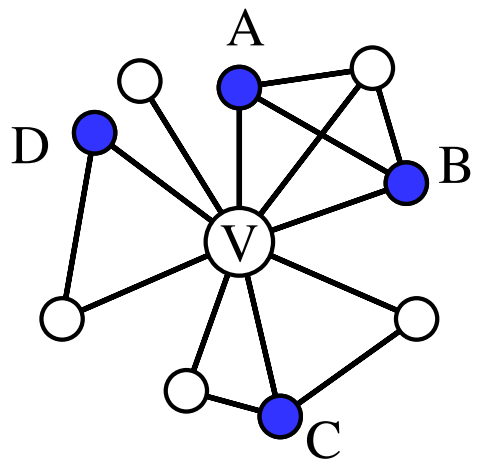




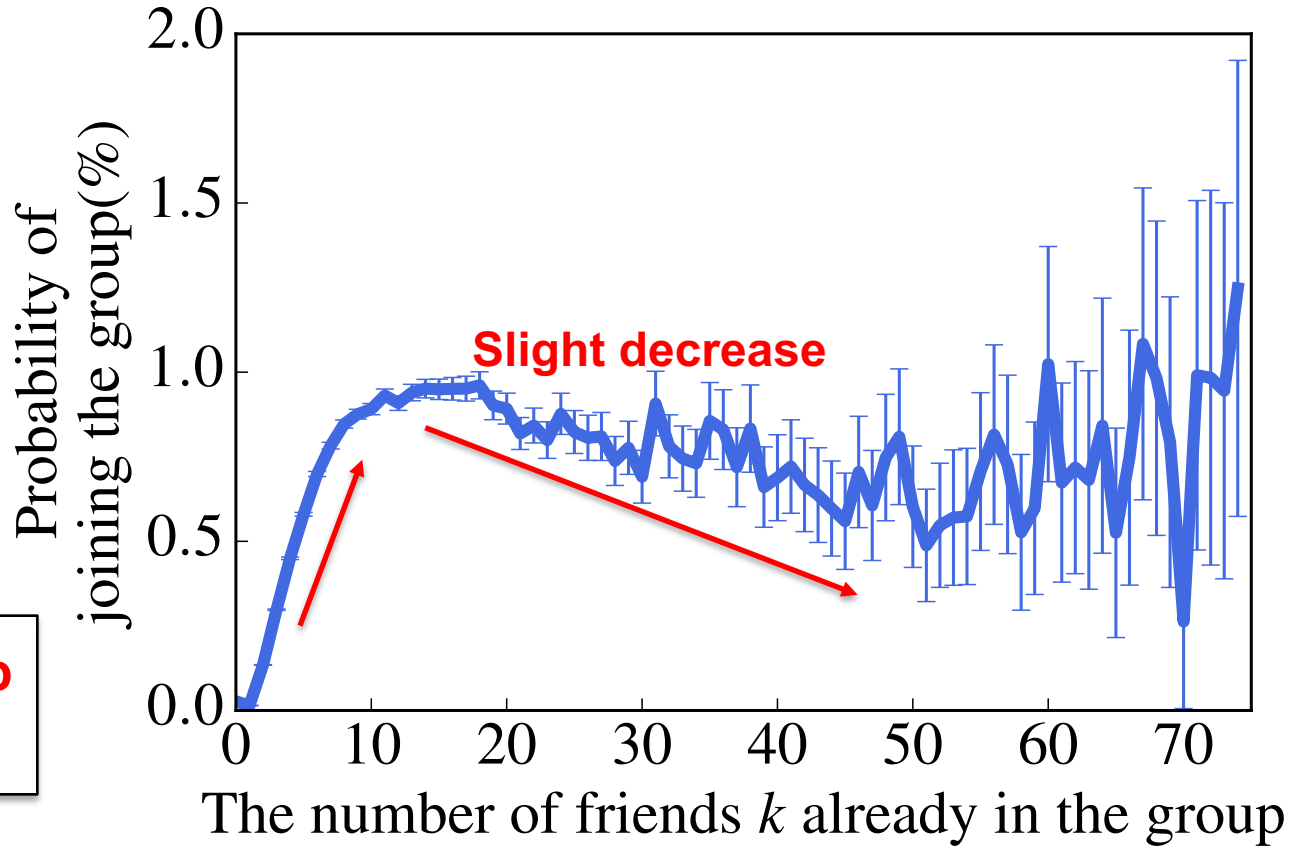
~80% of the first invitations happen within 5 days after the inviter joining the group

>80% of consecutive invitations by the same inviter happen within 2 days of interval

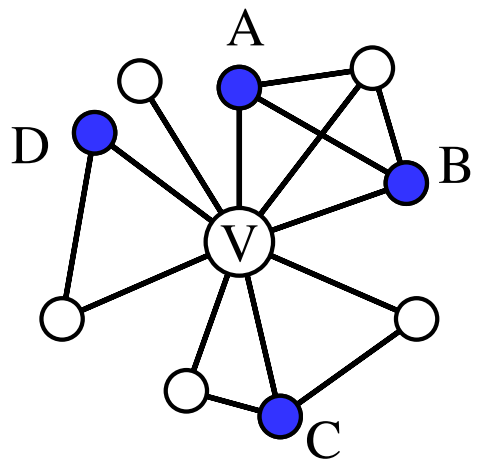
Membership Cascade—Invitees' Local Structure



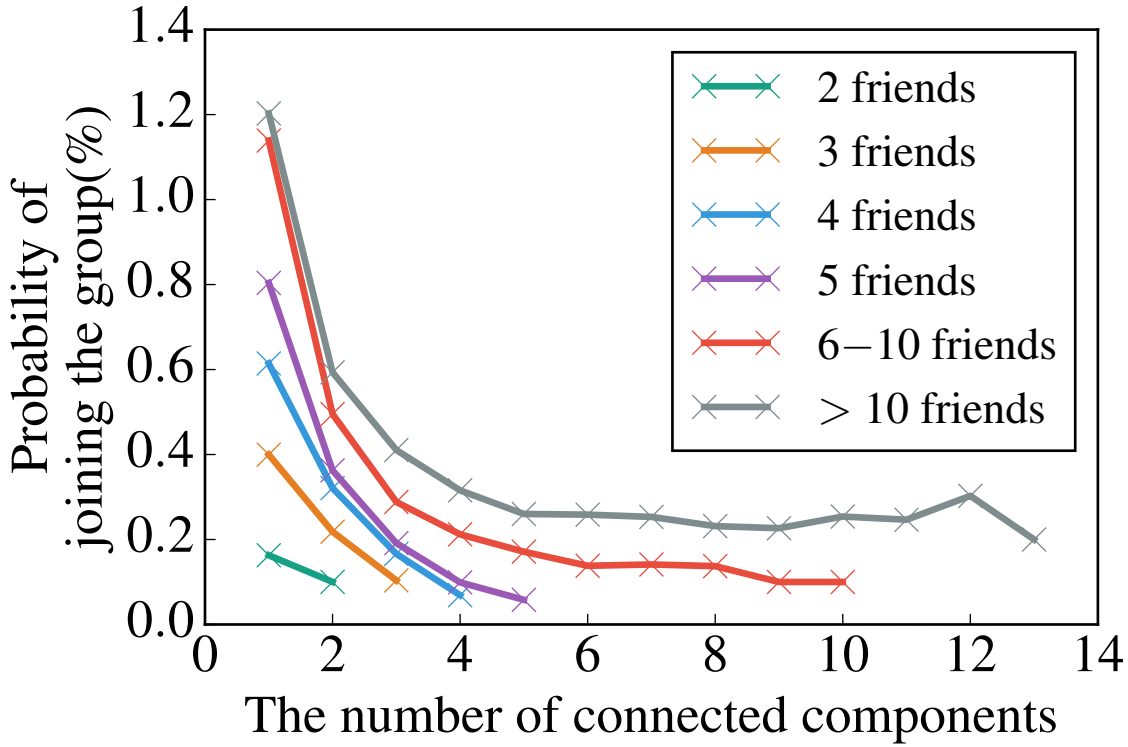
- V has 4 friends already in the group
- k=4



Membership Cascade—Invitees' Local Structure



- **V has 4 friends already in the group**
- **They form 3 connected components**



Zhang, Liu, Tang et al, IJCAI'2013

[1] Jing Zhang, Biao Liu, Jie Tang, Ting Chen, and Juanzi Li. Social Influence Locality for Modeling Retweeting Behaviors. In Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI'13)

Membership Cascade—Features

Inviter Level (for member u in group C at time T)

History Behavior	How long has it been since u invited others to C.
Local Structure	The number and fraction of u's friends in the group

Invitee Level (for user u in the fringe of group C at time T)

Demographics	User u's stated gender.
Local Structure	Number of friends already in the group.

Membership Cascade—Prediction

SVM 10-fold Cross Validation

Task	Feature Used	AUC	Precision	Recall	F1
Inviter	All	95.31	85.95	88.39	87.15
	-History Behavior	91.52	82.07	84.31	83.17
	-Local Structure	93.22	84.50	87.04	85.75
Invitee	All	98.66	54.55	93.47	68.69
	-Demographics	98.05	45.76	94.68	61.70
	-Local Structure	89.29	11.85	76.53	20.52

Summary

- We take the first step to study social messaging groups.
- We discover a strong dichotomy of groups in terms of their lifecycle.
- We define the membership cascade process and develop a model to predict the dynamics of the process.

Furture Research

- Coevolution of chat groups
- Comparison between information diffusion and membership cascade process.
- Role of chat group in the whole WeChat ecosystem

Thank you !

Collaborators:

Yixuan Li, John E. Hopcroft (**Cornell**)

Jie Tang (**THU**)

Qiang Yang (**HKUST**)

Zheng Lu, Hao Ye, Bo Chen (**Tencent**)