# A First Look at Developers' Live Chat on Gitter
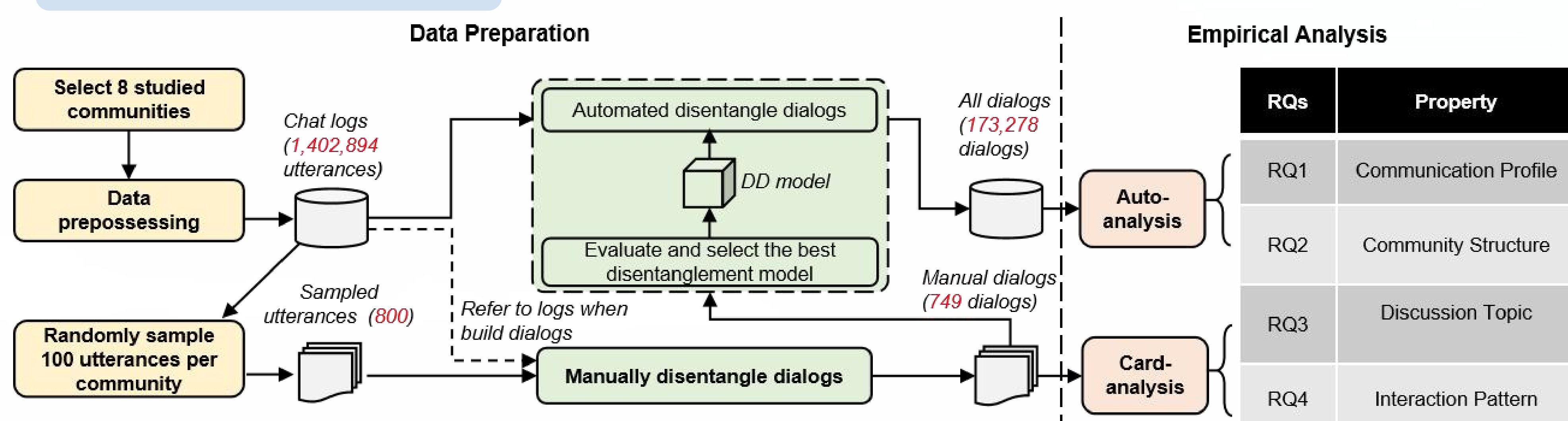## 基于在线聊天的开源社区特征分析

石琳 陈啸 杨叶 蒋涵芝 江子攸 牛楠 王青

the 29th ACM Joint European Software Engineering Conference and
Symposium on the Foundations of Software Engineering
(ESEC/FSE 2021)
主要联系人：石琳（15001193593，shilin@iscas.ac.cn）

## Background

Modern communication platforms such as Gitter and Slack play an increasingly critical role in supporting software teamwork, especially in open source development. Conversations on such platforms often contain intensive, valuable information that may be used for better understanding OSS developer communication and collaboration. However, little work has been done in this regard. To bridge the gap, this paper reports a first comprehensive empirical study on developers' live chat, investigating when they interact, what community structures look like, which topics are discussed, and how they interact.

## Method



**Data Preparation**

- Select 8 studied communities
- Data prepossessing
- Chat logs (1,402,894 utterances)
- Automated disentangle dialogs
- DD model
- Evaluate and select the best disentanglement model
- Randomly sample 100 utterances per community
- Sampled utterances (800)
- Refer to logs when build dialogs
- Manually disentangle dialogs
- All dialogs (173,278 dialogs)
- Manual dialogs (749 dialogs)

**Empirical Analysis**

| RQs | Property |
|---|---|
| RQ1 | Communication Profile |
| RQ2 | Community Structure |
| RQ3 | Discussion Topic |
| RQ4 | Interaction Pattern |

- Auto-analysis
- Card-analysis

RQ1 (Communication Profile): Do Gitter communities demonstrate consistent community communication profiles?
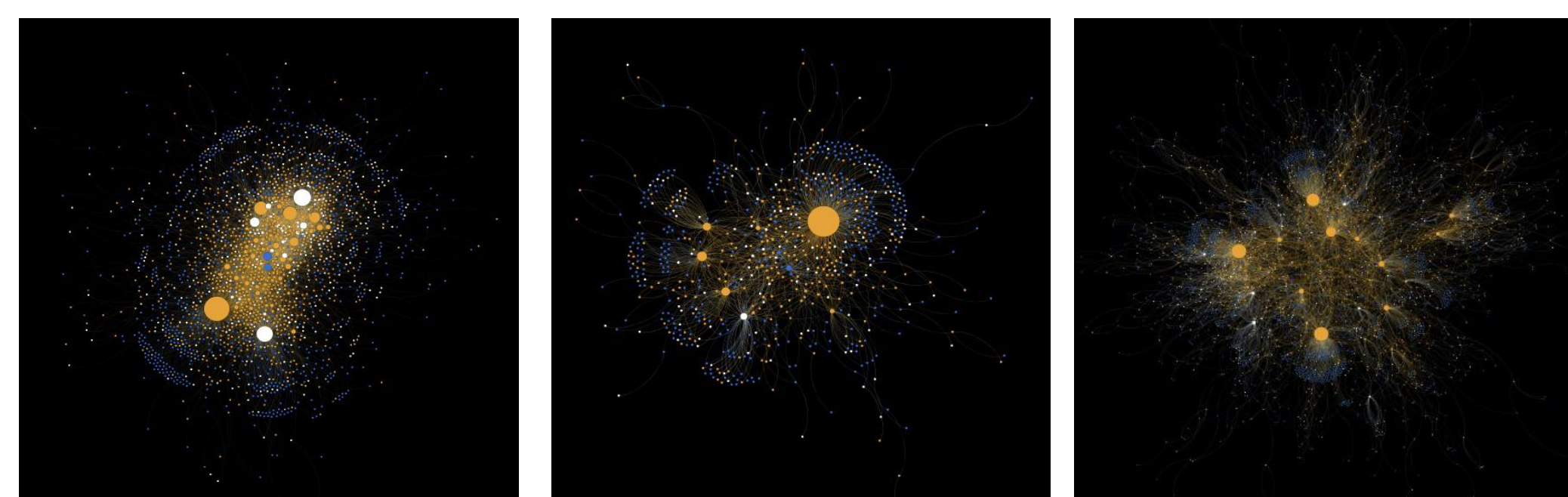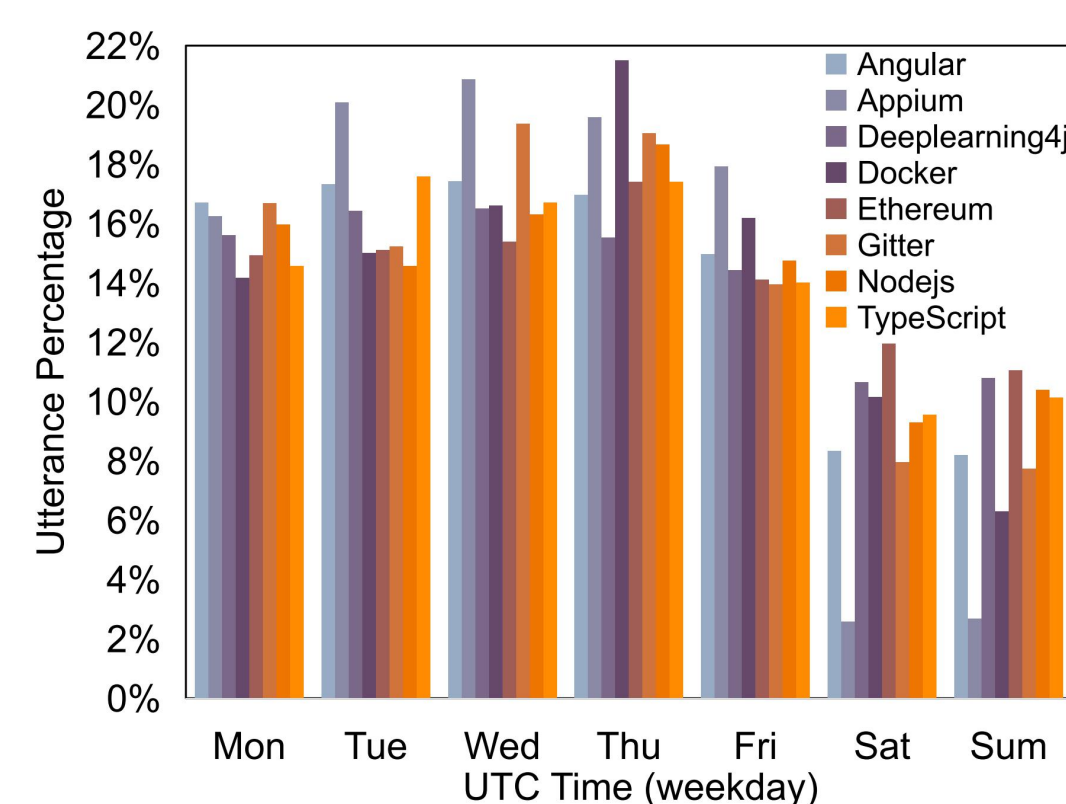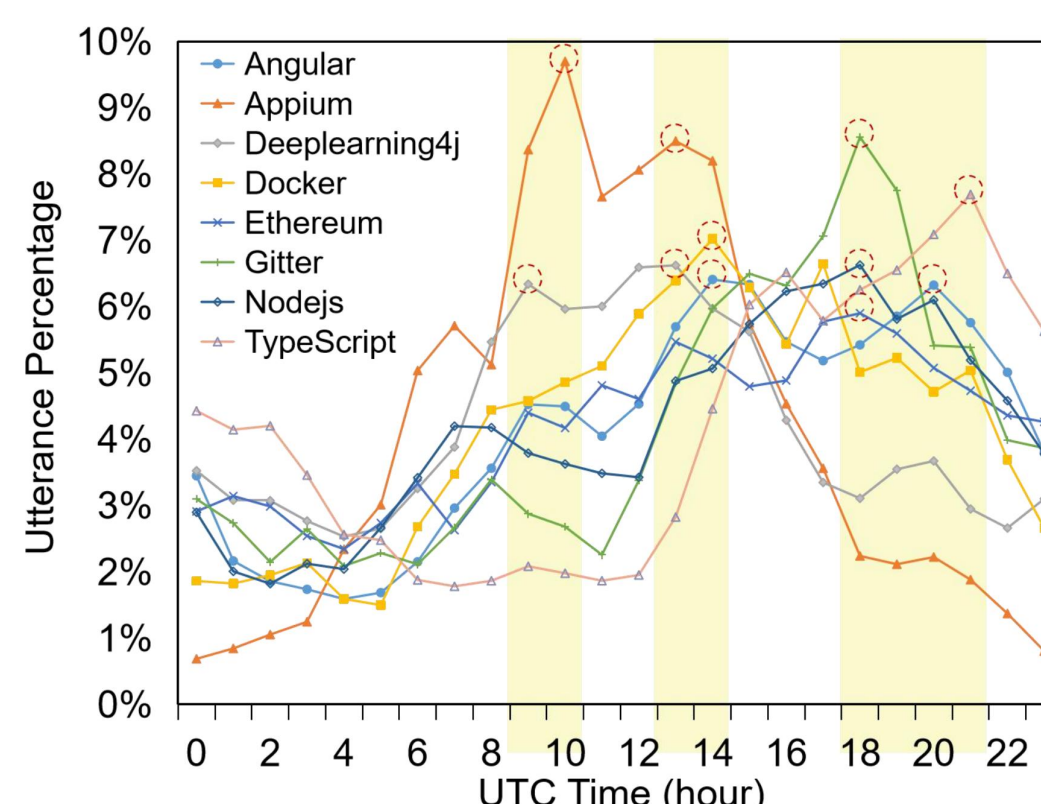RQ2 (Community Structure): What are the structural characteristics of social networks built from developer live chat data?
RQ3 (Dialog Topic): What are the primary topic types frequently discussed by developers in live chat?
RQ4 (Interaction Pattern): How do developers typically interact with each other in live chat?

## Result

**Answering RQ1:** The peak hours for live chat are from UTC 9 to 10, 13 to 14, and 18 to 21, while UTC 1 to 6 is the low-active hours. Developers are more likely to chat on workdays than weekends, especially on Wednesdays and Thursdays (UTC). Moreover, live chat gets 50% faster replies than the fast answers in Stack Overflow.





**Answering RQ2:** By visualizing the social networks of eight studied communities, we identify three social network structures for developers' live chat. Half of the communities (4/8) are Constellation networks. A minority of the communities (3/8) are Polaris networks. Only one community belongs to the Galaxy network.

Constellation    Polaris    Galaxy

**Answering RQ3:** Developers launch solution-oriented dialogs and problem-oriented dialogs more than knowledge-oriented dialogs. Nearly 1/3 of dialogs are about API usage. Developers discuss more error, unwanted behavior, and do-not-work, than reliability issues, performance issues, and test/build failures.
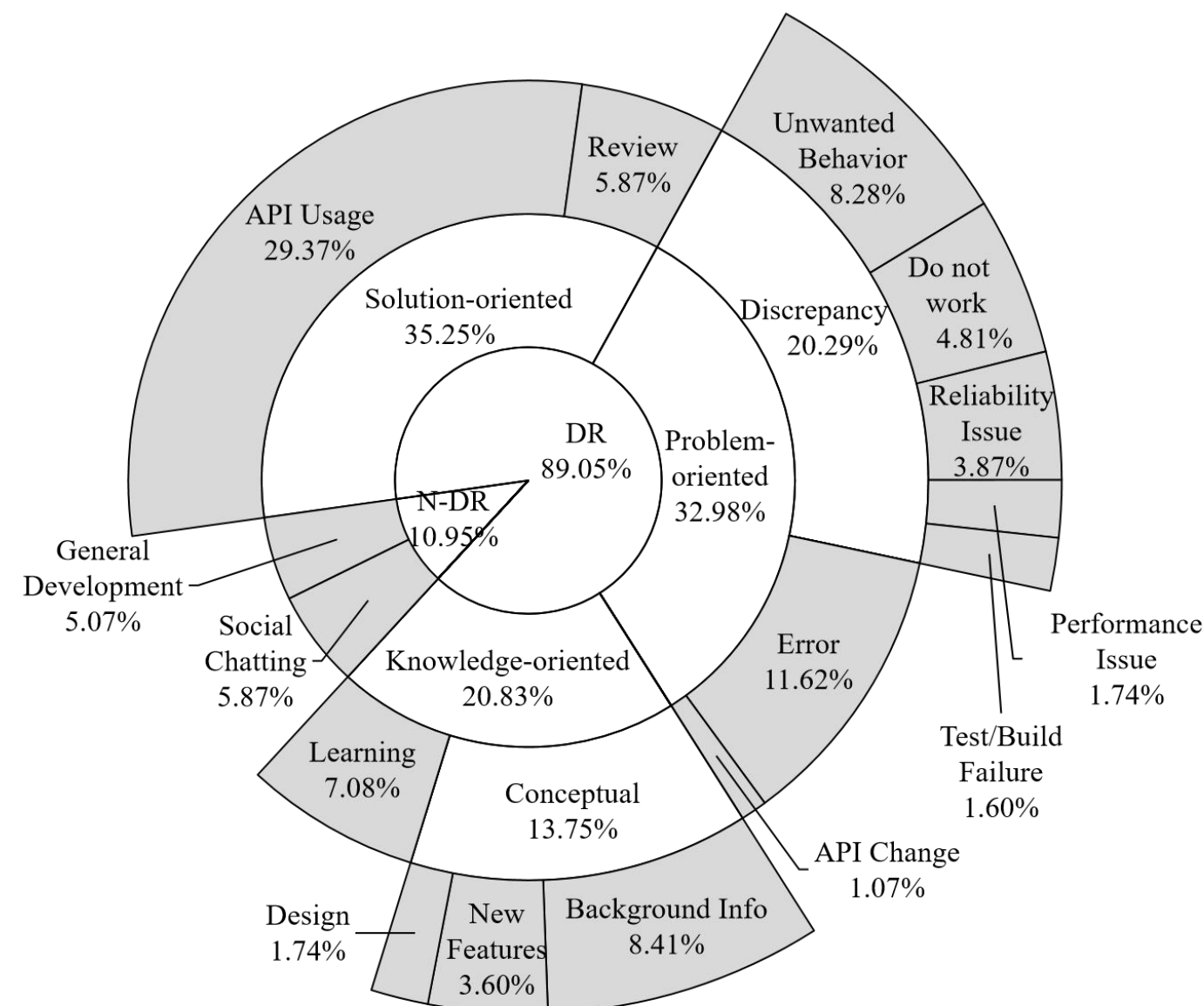


**Answering RQ4:** Six interaction patterns are identified in live chat: exploring solutions, clarifying answer, clarifying question, direct/discussed answer, self-answered monologue, and unanswered monologue.



P1: Multiple Solutions   P2: Clarifying Answers   P3: Clarifying Questions

P4: Direct/Discussed Answer   P5: Self-answered Monologue   P6: Unanswered Monologue
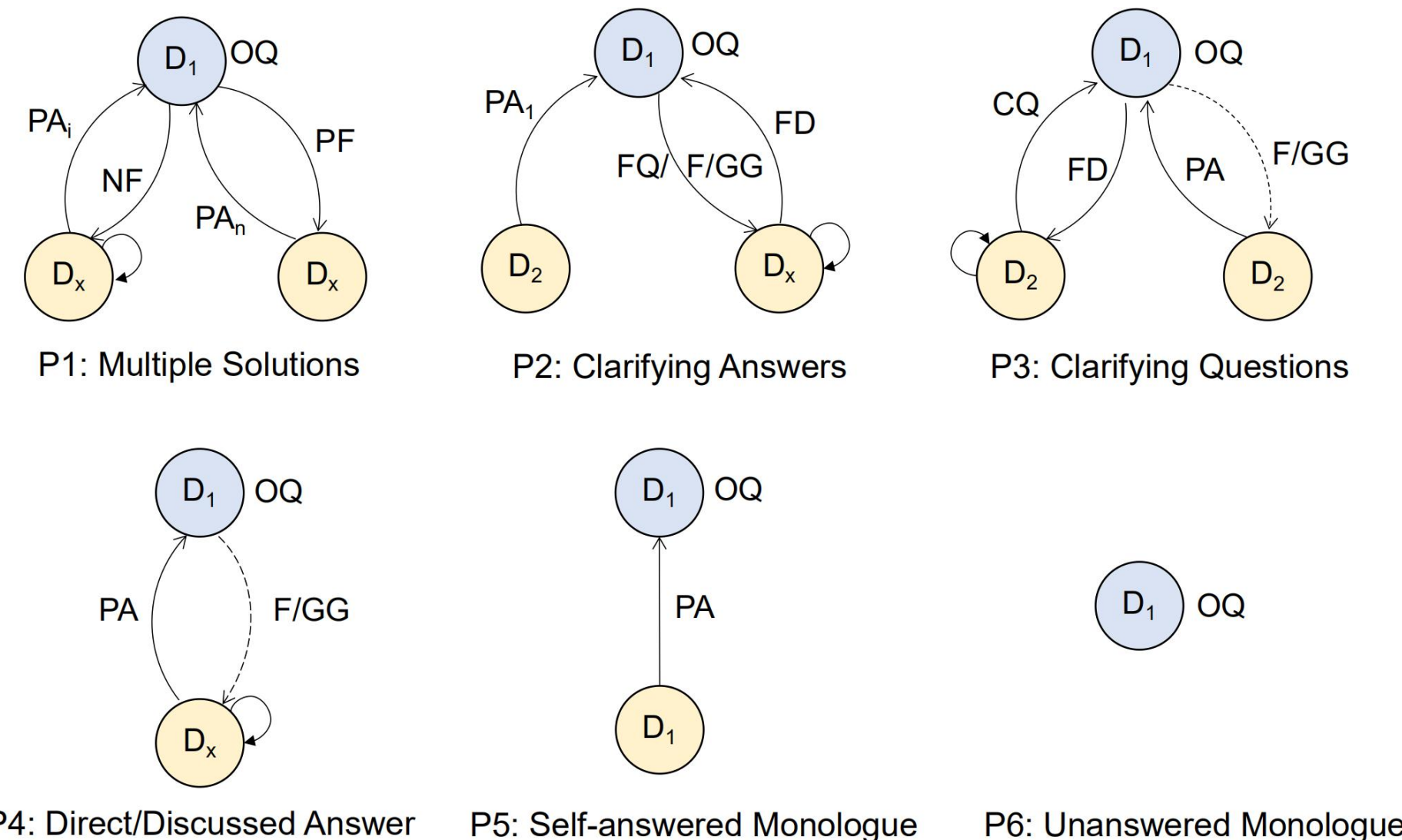
## Conclusion

- We have presented the first large-scale study to gain an empirical understanding of OSS developers' live chat.
- Based on 173,278 dialogs taken from eight popular communities on Gitter, we explore the temporal communication profiles of developers, the social networks and their properties towards the community, the taxonomy of discussion topics, and the interaction patterns in live chat.
- We provide recommendations for both OSS developers and communities, highlight advanced features for online communication platform vendors, and provoke insightful future research questions for OSS researchers.