

# Studying Forum Dynamics from a User Engagement Perspective

[Extended Abstract]

Andrey Kan  
NICTA Victoria  
Research Laboratory,  
Department of CSSE,  
University of Melbourne,  
Australia  
akan@csse.unimelb.edu.au

Conor Hayes  
Digital Enterprise  
Research Institute,  
National University of Ireland,  
Galway, Ireland  
conor.hayes@deri.org

Jeffrey Chan  
Digital Enterprise  
Research Institute,  
National University of Ireland,  
Galway, Ireland  
jkc.chan@deri.org

Bernie Hogan  
Oxford Internet Institute,  
University of Oxford  
United Kingdom  
bernie.hogan@oii.ox.ac.uk

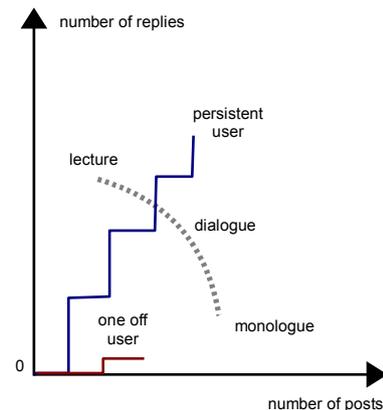
Online forums are popular websites that provide extensive communication records of users over time. Systematic analysis of these records can facilitate the mathematical modeling of communication and the development of forum management tools. For example, the models can be capable of prediction whether a user is likely to elicit responses in the future, based on the past observations.

Many researchers have tried different approaches to forum data analysis, such as a focus on user roles, discussion diversity and growth dynamics. For example, Both Viegas and Smith [4] and Welser et al. [5] use visualization to derive structural features of the roles of individual users thus indicating clear differences between consistent “answer people” and those preferring discussions. Chan and Hayes [1] extend this work by categorizing such roles and exploring the composition of forums in terms of a diversity of roles. Kumar et al. [3] and Gonzalez-Bailon et al., [2] model growing trees of discussion threads.

This previous work has proven germane, but is constrained in several ways. First, these studies tend to assume that users exhibit consistent behavior over time. Second, studies that look to temporal patterns over time instead of assuming consistency (such as *AuthorLines* [4]) tend to be descriptive, visual and difficult to scale.

To address these limitations, we propose a method to represent a user’s engagement in communication over time as a *user path* that is a line in a space of user features. In this paper, we focus on two features – the number of posts made

and the number of replies received. In this case, a *user path* is a line in the 2D space. The path starts at point (0, 0) and whenever a user makes a post the path moves one unit right. Whenever the user receives a reply the path moves up one unit. The path can be interpreted as shown in Fig. 1.



**Figure 1: Solid lines represent user paths. The longer path corresponds a more persistent user, i.e., a user that have a regular activity in the forum. The average slope of the path is related to the reciprocity of communication. For example, a path closer to x-axis contains longer consecutive posting sequences. Such a path can be classified as a “monologue”. In contrast, a “dialogue” is a more balanced path where posts tend to be followed by replies.**

Our method allows us to study the following research questions that were previously unaddressed.

1. To what extent do users exhibit consistency in communication patterns over time?

2. Can we compare and contrast the evolution of multiple users across forums?
3. Can we numerically describe forums in terms of temporal behavior of users?

To address the first question, we fit user paths with straight line. We find that the user path can fit with straight line reasonably well. This indicates that users indeed tend to exhibit consistent communication patterns over time. At the same time users can have variations from the straight line. We measure these variations, and thus the extent of consistency, with the root mean square error (RMSE) of the fit.

To address the second question, we superimpose paths for different users in a forum (Fig. 2). We find some properties common to all users. Specifically we find that there are dead zones in the feature space (i.e., areas where the probability of observing a user is low). A user found in these zones exhibits abnormal behavior with respect to the majority of users. Such a user, can be, for example, a spammer, and thus users with abnormal behavior might be of interest for forum administrators.

To address our last research question, we characterize forums with macro features, such as the mean length and the mean slope for the set of user paths. We validate the macro features with hierarchical clustering of forums (Fig. 3).

In summary, the main contributions of our work are as follows.

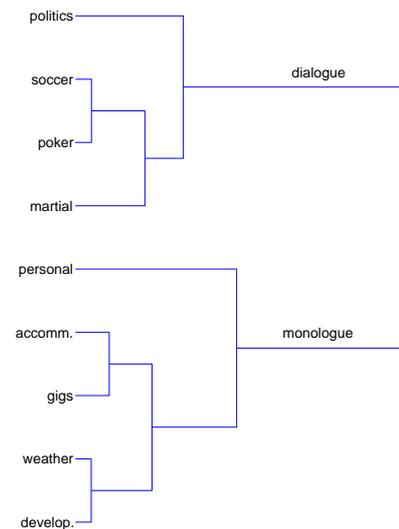
- A novel method for studying temporal evolution of individual users, measuring their behavior consistency, and scaling the method up to the levels of a forum and a forum board.
- Found patterns of consistent user engagement and dead zones.
- A method for an automated description and categorization of forums in terms of evolving user patterns.

In addition to our main results, we emphasize the flexibility and generality of our approach, since it can be applied to other domains, such as email or phone communications. Here the numbers of posts and replies can be replaced by the numbers of sent and received emails or outgoing and incoming calls respectively. Further, one can consider a user path in different feature spaces, for example, with posts and replies weighted as they are voted on by users, in cyberspaces such as Reddit and YouTube. Finally, our analysis can be performed on the collective evolution of communities rather than individual users, revealing community level patterns.

We suggest the following applications for our results: (i) predict whether a user is likely to be involved in conversations in future, based on past observations; (ii) identify users whose behavior significantly differs from the normative behavior for a given forum; (iii) categorize forums by their ability to elicit responses and apply different management policies for different categories of forums.

## Acknowledgments

This work is partially supported by Science Foundation Ireland (SFI) under CLIQUE Strategic Cluster, grant number 08/SRC/I1407. This work is partially supported by National ICT Australia (NICTA). NICTA is founded by the



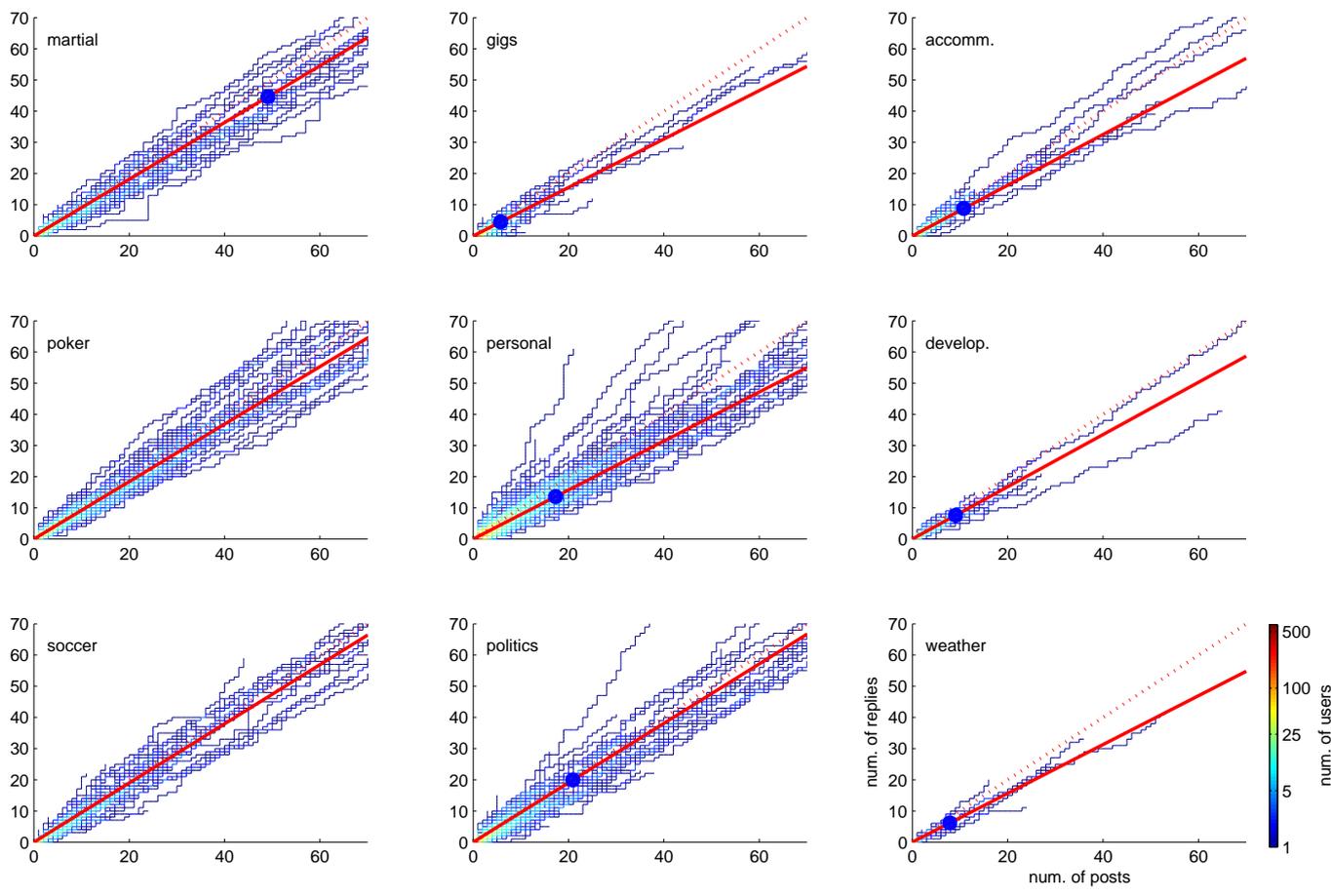
**Figure 3: Hierarchical clustering of the forums based on macro features that characterize forum appearance. There are two prominent groups of forums. We label the first group as “dialogue”, because the forums in this group have relatively high reciprocity of communication (the average forum slope is closer to the 45 degree line). We label the second group as “monologue”, because of a lower reciprocity of communication (the average forum slope is closer to the x-axis).**

Australian Government’s Backing Australia’s Ability initiative, in part through the Australian Research Council.

Andrey Kan has conducted this work while at Digital Enterprise Research Institute, National University of Ireland, Galway.

## 1. REFERENCES

- [1] J. Chan and C. Hayes. Decomposing Discussion Forums using User Roles. In *Decomposing Discussion Forums using User Roles*, *Proceedings of the Second Web Science Conference*, 2010.
- [2] S. Gonzalez-Bailon, A. Kaltenbrunner, and R.E. Banchs. The structure of political discussion networks: a model for the analysis of online deliberation. *Journal of Information Technology*, 25(2):230–243, 2010.
- [3] R. Kumar, M. Mahdian, and M. McGlohon. Dynamics of conversations. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 553–562. ACM, 2010.
- [4] F.B. Viégas and M. Smith. Newsgroup crowds and authorlines: Visualizing the activity of individuals in conversational cyberspaces. In *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*, page 10. IEEE, 2004.
- [5] H.T. Welser, E. Gleave, D. Fisher, and M. Smith. Visualizing the signatures of social roles in online discussion groups. *Journal of Social Structure*, 8(2), 2007.



**Figure 2: Visualization of nine forums from Boards.ie.** For each forum, a random sample of all users registered in year 2006 was selected. Each user was tracked until January 2008. Each line is a path of one user. User paths can overlap, and color shows how many paths overlap at different points. Dashed red line is the 45 degree line, solid red line shows the average slope of the forum. Blue circle shows the average length of user path for the forum (for “soccer” and “poker” forums the average length is above 50).