Dan Cosley: Research statement

My goal is to improve research and design in social media systems by better integrating technical activities such as behavior modeling, interface design, and system building with understanding of social dynamics as expressed by social science theories. In particular, I am inspired by Will Hill et al.'s idea of edit wear and read wear [21] that people leave traces of their activity through interacting with digital objects, which foreshadows today's practical and research interest in "big data". Much of my work revolves how to derive meaning from these traces, and I see three broad approaches to this meaning-making in both my work and in others'. In one camp, meaning-making resides in the system, which uses the data to make interferences about users and personalize interfaces. In the second camp, researchers use the data to make meaning in the form of theories and models of behavior. In the third camp, people themselves reflect on data to make meaning about themselves, their relationships, and their communities. Below, I'll talk about each of these camps and my contributions toward them, then close with research I am pursuing to bring these camps together.

Systems making meaning: user modeling and personalization

One reason digital traces are so compelling is that they give information about people's interests, knowledge, and preferences. This information has been used to great effect by collaborative filtering-based recommender systems [39], which use data such as rating movies, reading news articles, or purchasing items to recommend new movies, articles, or purchases. When I started my PhD at Minnesota's GroupLens lab, much of the algorithmic work had already been done. Thus, my early work focused on **making recommender systems more useful**, including recommending for groups [33], supporting new users [38], evaluating systems based on online usage rather than offline datasets [14], and recommending in non-entertainment domains such as research papers [14,29].

This was also where I first adopted the goal of bringing system design and social science theory together, as part of the CommunityLab collaboration between Minnesota, Michigan, and CMU. This can be done in a primarily analytical way: theories about conformity [2] and the idea of computers as social actors [31] suggest showing predicted ratings in a recommender system might influence people's own ratings. This is true, and has negative consequences both for the systems themselves and studies based on data collected by these systems such as the MovieLens datasets [13]. It can also be done in a more design-focused way: **theories of motivation can be used to derive design strategies to encourage people to contribute to the community**. These including making the value of one's contributions clear [26,27] and reducing the cost of contributing through intelligent task routing, or matching people with appropriate work [11]. My implementation of task routing in SuggestBot, an article suggestion system in Wikipedia [12], has been ported to several Wikipedia languages and has been applied to elements of Wikipedia including the community portal and the Teahouse welcoming page for newcomers.

One way to think of the systems camp is that it treats trace data as the key to a locked door. User modeling unlocks the door, allowing the owner of the system to use this knowledge to enter our minds and affect our behavior (not unlike the idea of persuasive interfaces). This often

benefits the owner in e-commerce scenarios; it may benefit communities as described above; and may also help individuals by reducing their search costs or increasing their happiness. But the person who generated the data is rarely an active agent in the process of understanding, and the soul of this work is in the machine.

Researchers making meaning: computational social science

Another view of these data is as a large, but noisy, aggregate record that can be used to test or generate theories and models of social behavior, aka "computational social science" [24]. My development of the SuggestBot article recommender in Wikipedia [12] led me to understand it as both a dataset and a community, and **thinking of Wikipedia collaboration as a network of authors and articles** led to a number of NSF-funded projects with folks from communication, computer science, information science, and sociology.

These included exploring how theories of social influence affect the diffusion of tools in Wikipedia [56,57], inferring people's roles based on the activities they engage in [54,55], understanding discourse patterns in discussions around article production [5,6,48], and modeling article quality [51]. The most satisfying part of this work **combined large-scale modeling of behavior with close analysis of interactions** to understand how homophily of interests and social influence interact and predict when people will edit articles or talk with other editors [17]. Models, methods, and insights from this work has been taken up by many studies of influence and diffusion in social networks (e.g., [48]), in the design of recommendation algorithms (e.g., [22]), and in understanding collaboration in Wikipedia (e.g., [32]). In particular, most collaboration is planned: people first talk about how to work together, then carry out the tasks. In distributed, decentralized, loosely structured collaborations like Wikipedia, however, the process is often reversed: people work independently on the same articles and start talking when they discover a need to coordinate.

The work above was primarily analytical, but **system-building can also contribute toward understanding behavior**. For instance, most tagging systems are designed with the idea that tags support categorization and search; however, early on we [41] and others [1,52] noticed that labeling is only part of the story. With communication PhD student Jenn Thom and several undergraduates in Geri Gay's lab, I designed and deployed tagging systems that de-emphasize categorization and demonstrated that people use tags for a variety of cognitive and expressive purposes. The Artlinks and MobiTags museum tagging systems were designed to meet social and experiential goals of museum-going that are commonly neglected in informational tools [4]. Patterns of use and interview results showed that people used tags not just to find items, but to think about both items and the people who tagged them [10,15] to signal their expertise [50], and to express mood and humor [7,8].

In this camp, big data is a window through which researchers observe human behavior. As with the system camp, this may have side benefits—these understandings might inform designs work, while systems designed to address research questions may support new needs. And, as with the system camp, the people who generate the data are typically objects of study, not agents of understanding; the soul of this research is in the hypotheses.

People making meaning: reflecting on data

Hill et al.'s original vision involved helping workgroups better understand and coordinate their behavior by presenting trace data in the interface [21]. This idea of presenting trace data back to its creators, as a kind of mirror, is an increasingly common theme in HCI, around personal informatics [25], lifelogging [20], and visualizing interaction [19,58]. These systems are often critiqued for having minimal theoretical guidance and unclear goals [42]; my aim is to **design to meet real needs and align with existing practices**.

Thus far I have focused on supporting individual reminiscence and reflection. These are important psychological processes for people of all ages [53] and a number of systems have attempted to support social reminiscing in families [37] around artifacts such as pictures, mementoes, and sounds. My insight was that the **data people accumulate in social media could support personal reflection** as well as social reminiscence. Supported by my CAREER grant, I built the Pensieve system [9], which emails people content they had previously created and provides a convenient diary interface to help people think and write about the past. A large-scale deployment led by research intern Tejaswi Peesapati and PhD student Victoria Schwanda Sosik showed that Pensieve supports common goals of reminiscing [35], while follow-on work showed the need for presenting culturally congruent stimuli for reminiscing and the value of place as a trigger for reminiscing [34,36]. This is a growth area in HCI—there was a strong response to a CHI 2011 workshop I led—and its design ideas and focus on evaluation directly informed commercial systems such as Timehop (for which Jon Baxter, an undergrad lead researcher on Pensieve, is a senior engineer) and research systems such as Echo [22].

More recent work with Victoria and PhD student Xuan Zhao has shifted toward social uses of trace data, influenced by their interests and an instructive failure to adapt Pensieve to support social reminiscing [16]. In particular, we have found that choices around creating, consuming, and deleting information in Facebook are complicated because **people simultaneously manage concerns around identity management, relationship dynamics, visibility of activity to third parties, and self-archiving**. However, they are also skilled at appropriating interface features to manage these tensions, creating spaces for privacy and separation of concerns (with some uncertainty because of the opaqueness of Facebook's privacy mechanisms) [47,59,60].

Impact and future work

The research community values my work in each of these areas. It appears in high-quality venues such as CSCW and CHI, has won three best paper awards, has been cited over 3,100 times, and appears in a number of course syllabi. My work on Wikipedia led to my being technical chair of the 2012 WikiSym conference, while my body of work in online communities and social media was recognized with my appointment as general co-chair of CSCW 2015. I've collaborated with dozens of students and faculty at all levels, and PhD students I've worked closely with have gone on to careers in both academic and industrial research. SuggestBot and Pensieve formed the basis for my NSF CAREER grant, my work around contribution to online communities was a big piece of a successful NSF HCC medium grant I helped write as a PhD student, and my work around modeling behavior in Wikipedia contributed to an HCC large

grant here at Cornell. Finally, the SuggestBot and Pensieve systems continue to have active user communities years after their development.

I expect this impact to continue. Based on the growth of grant funding, research, and companies around big data, questions around extracting meaning from traces of online activity will continue to be important for both research and practice. Having done work in all three camps, I think **the big payoff will come around work in which systems, researchers, and people work together to make meaning out of big data**. Below I'll sketch two ongoing streams of my work that bring the camps together.

The first stream is rooted in computer science PhD student Amit Sharma's idea of network-centric recommendation [46]. Other research has started to incorporate social network information into recommendation algorithms to support individual choices [28], but Amit's asking **"What would a recommender system designed to be social from the ground up look like?"** is a fundamental rethinking that begs for a marriage of sociology and design. Metrics and algorithms might encourage social experiences, interpersonal recommendation sharing, and diffusion of items across communities [43] in addition to accuracy. Understanding how information diffusion and homophily shape preferences can inform algorithms, helping systems with only local information do as well as those with access to the full network in some cases [45]. Knowing how people make sense of social explanations such as "Dan Cosley and two friends liked this research statement" can help systems increase people's willingness to try new items and perhaps their ability to predict which items they will like [44]. There are dozens of questions at the intersection of social behavior, information sharing, and recommendation that will have both research and practical import.

The second strand revolves around the insight that some social media content is particularly meaningful for reflection [16]. The $64,000 question here is **how can we model whether content is meaningful given the content and people's context and goals, and use those models to support psychological well-being?** This aligns with recent work around modeling mental health from social media data [18,30], but with more focus on both theoretical grounding and technical interventions. Victoria's dissertation addresses how social media content and platforms can help people engage with positive psychology interventions [40] aimed at people at all levels of well-being. My work with Natalie Bazarova on self-disclosure in social media reveals that interface elements, perceived and actual audiences, and personal goals and characteristics all affect people's decisions about what to disclose when, to whom [3], meaning that effective models of online disclosure will require addressing multiple theories and levels of analysis. Our ongoing work focuses on collecting data from a variety of media and from people at risk of mental help problems that we will use to build better models of online self-disclosure and inform systems that help people reflect on self-disclosure behavior and get and give support. Both the domain of well-being and the general question of finding meaningful content in the sea of big data are critically important questions.

Thanks for taking the time to read this and feel free to contact me if you have any questions at danco@cs.cornell.edu.

References

Conference acceptance rates are presented where known for my papers.

1. Ames, M., & Naaman, M. (2007, April). Why we tag: motivations for annotation in mobile and online media. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 971-980.
2. Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. *Groups, Leadership, and Men*, 177–190.
3. Bazarova, N. N., Taft, J., Choi, Y., Cosley, D. (2013). Managing impressions and relationships on Facebook: Self-presentational and relational concerns revealed through the analysis of language style. *Journal of Language and Social Psychology*.
4. Bell, G. (2002). Making Sense of Museum: The Museum as 'Cultural Ecology': A study. *CIMI whitepaper, Intel Corporation*.
5. Black, L. W., Welser, H. T., DeGroot, J. M., Cosley, D. (2008). `Wikipedia is not a democracy': Deliberation and Policy-Making in an Online Community. *International Communication Association*.
6. Black, L. W., Welser, H. T., DeGroot, J. M., Cosley, D. (2011). Self-governance through Group Discussion: Wikipedia's Policy Making Implications for Virtual Teams. *Small Group Research*.
7. Cheng, J., Cosley, D. (2010). kultagg: ludic design for tagging interfaces. *Proc. GROUP 2010*. (27%)
8. Cheng, J, Cosley, D. (2013). How annotation styles influence content and preferences. *Proceedings of HyperText 2013*. (17%)
9. Cosley, D., Akey, K., Alson, B., Baxter, J., Broomfield, M., Lee, S., Sarabu, C. (2009). Using Technologies to Support Reminscence. *BCS HCI 2009*. Cambridge, UK.
10. Cosley, D., Baxter, J., Lee, S., Alson, B., Adams, P., Nomura, S., Sarabu, C., Gay, G. (2009). MobiTags: Supporting Semantic, Spatial, and Social Interaction in Museum Spaces. *CHI 2009*. (25%)
11. Cosley, D., Frankowski, D., Terveen, L., Riedl, J. (2006). Using Intelligent Task Routing and Contribution Review to Help Communities Build Artifacts of Lasting Value. *CHI 2006*. (24%)
12. Cosley, D., Frankowski, D., Terveen, L., Riedl, J. (2007). SuggestBot: Using Intelligent Task Routing to Help People Find Work in Wikipedia. *IUI 2007*. (22%)
13. Cosley, D., Lam, S. K., Albert, I., Konstan, J., Riedl, J. (2003). Is Seeing Believing? How Recommender Systems Influence Users' Opinions. *CHI 2003,* 585–592. (16%)
14. Cosley, D., Lawrence, S., Pennock, D. M. (2002). REFEREE: An open framework for practical testing of recommender systems using ResearchIndex. *VLDB 2002*, 35–46. (16%)
15. Cosley, D., Lewenstein, J., Herman, A., Holloway, J., Baxter, J., Nomura, S., Boehner, K., Gay, G. (2008). ArtLinks: Fostering Social Awareness and Reflection in Museums. *CHI 2008*. (22%)
16. Cosley, D., Sosik, V. S., Schultz, J., Peesapati, S. T., Lee, S. (2012). Experiences with designing tools for everyday reminiscing. *HCI*.
17. Crandall, D., Cosley, D., Huttenlocher, D., Kleinberg, J., Suri, S. (2008). Feedback effects between similarity and social influence in online communities. *KDD 2008*. (10%)
18. De Choudhury, M., Counts, S., & Horvitz, E. (2013). Predicting postpartum changes in emotion and behavior via social media. In *Proceedings of the 2013 ACM annual conference on Human factors in computing systems*, 3267–3276.
19. Donath, J., Karahalios, K., Viégas,F. (1999). Visualizing Conversation. *Journal of Computer-Mediated Communication.* 4(4).
20. Gemmell, J., Bell, G., Lueder, R. (2006). MyLifeBits: a personal database for everything.

Communications of the ACM 49(1), 88–95.

21. Hill, W. C., Hollan, J. D., Wroblewski, D., McCandless, T. (1992). Edit wear and read wear. *Proceedings of CHI 1992*, 3–9.
22. Isaacs, E., Konrad, A., Walendowski, A., Lennig, T., Hollis, V., Whittaker, S. (2013). Echoes from the past: how technology mediated reflection improves well-being. In *Proc. SIGCHI,* 1071–1080.
23. Jamali, M., Ester, M. (2009). TrustWalker: a random walk model for combining trust-based and item-based recommendation. In *Proc. SIGKDD*, 397–406.
24. Lazer, et al. (2009). Computational Social Science. *Science* 323, 721–723.
25. Li, I., Dey, A., Forlizzi, J. (2010). A stage-based model of personal informatics systems. *Proc. CHI 2010*, 557–566.
26. Ling, K., Beenen, G., Ludford, P., Wang, X., Chang, K., Li, X., Cosley, D., Frankowski, D., Terveen, L., Rashid, A. M., Resnick, P., Kraut, R. (2005). Using social psychology to motivate contributions to online communities. *J. Computer-Mediated Communication*, 10(4).
27. Ludford, P., Cosley, D., Frankowski, D., Terveen, L. (2004). Think Different: Increasing Online Community Participation Using Uniqueness and Group Dissimilarity. *CHI 2004*, 631–638. (16%)
28. Ma, H., Zhou, D., Liu, C., Lyu, M. R., King, I. (2011). Recommender systems with social regularization. In *Proc. WSDM,* 287–296.
29. McNee, S., Albert, I., Cosley, D., Gopalkrishnan, P., Lam, S. K., Rashid, A. M., Konstan, J., Riedl, J. (2002). On the Recommending of Citations for Research Papers. In *Proceedings of CSCW2002*, 116–125. (20%)
30. Moreno, M. A., Jelenchick, L. A., Egan, K. G., Cox, E., Young, H., Gannon, K. E., & Becker, T. (2011). Feeling bad on Facebook: Depression disclosures by college students on a social networking site. *Depression and anxiety*, 28(6), 447–455.
31. Nass, C., Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 60(1):81–103.
32. Nemoto, K., Gloor, P., Laubacher, R. (2011). Social capital increases efficiency of collaboration among Wikipedia editors. In *Proc. Hypertext*, 231–240.
33. O'Connor, M., Cosley, D., Konstan, J. A., Riedl, J. (2001). PolyLens: A Recommender System for Groups of Users. *ECSCW 2001*, Bonn, Germany, 199–218. (19%)
34. Peesapati, S. T., Schwanda, V., Schultz, J., Cosley, D. (2010). Triggering memories with online maps. *Proceedings of ASIST 2010*. (30%)
35. Peesapati, S. T., Schwanda, V., Schultz, J., Lepage, M., Jeong, S., Cosley, D. (2010). Pensieve: Supporting Everyday Reminiscence. *CHI 2010*. (22%)
36. Peesapati, S. T., Wang, H-C., Cosley, D. (2010). Intercultural human-photo encounters: How cultural similarity affects perceiving and tagging photographs. *Proceedings of ACM International Conference on Intercultural Collaboration (ICIC 2010)*. (39%)
37. Petrelli, D., Villar, N., Kalnikaite, V., Dib, L., & Whittaker, S. (2010, April). FM radio: family interplay with sonic mementos. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* 2371–2380.
38. Rashid, A. M., Albert, I., Cosley, D., Lam, S. K., McNee, S., Konstan, J.A., Riedl, J. (2002). Getting to Know You: Learning New User Preferences in Recommender Systems. *IUI 2002*, 127–134. (33%)
39. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J. (1994). GroupLens: An open architecture for collaborative filtering of netnews. *Proceedings of CSCW 2004*, 175–186.

40. Seligman, M. E., Steen, T. A., Park, N., Peterson, C. (2005). Positive psychology progress: empirical validation of interventions. *American Psychologist*, 60(5), 410.

41. Sen, S., Lam, S. K., Rashid, A. M., Cosley, D., Frankowski, D., Osterhouse, J., Harper, F. M., Riedl, J. (2006). Tagging, communities, vocabulary, evolution. In *Proc. CSCW,* 181–190. (22%, best paper)

42. Sellen, A. J., Whittaker, S. (2010). Beyond total capture: a constructive critique of lifelogging. *Commun. ACM.* 53(5), 70–77.

43. Sharma, A., Cosley, D. (2011). Network-Centric Recommendation: Personalization with and in Social Networks. *Proceedings of SocialCom 2011.* (10%)

44. Sharma, A., Cosley, D. (2013). Do Social Explanations Work? Studying and Modeling the Effects of Social Explanations in Recommender Systems. *WWW 2013.* (15%)

45. Sharma, A., Gemici, M., Cosley, D. (2013). Friends, Strangers, and the Value of Ego Networks for Recommendation. *ICWSM 2013.*

46. Sharma, A., Malu, M., Cosley, D. (2011). PopCore: A system for Network-Centric Recommendations. In *Proc. 3rd Workshop on Recommender Systems and the Social Web.*

47. Sosik, V. S., Zhao, X., Cosley, D. (2012). See Friendship, Sort of: How Conversation and Digital Traces Might Support Reflection on Friendships. *CSCW 2012.* (39%).

48. Tang, J., Sun, J., Wang, C., & Yang, Z. (2009, June). Social influence analysis in large-scale networks. In *Proc. SIGKDD,* 807–816.

49. Thom-Santelli, J., Cosley, D., Gay, G. (2009). What's Mine is Mine: Territoriality in Collaborative Authoring. *CHI 2009.* (25%)

50. Thom-Santelli, J., Cosley, D., Gay, G. (2010). What Do You Know? Experts, Novices and Territoriality in Collaborative Systems. *CHI 2010.* (22%)

51. Warncke-Wang, M., Cosley, D., Riedl, J. (2013). Tell Me More: An Actionable Quality Model for Wikipedia. *WikiSym 2013.*

52. Wash, R., Rader, E. (2007). Public Bookmarks and Private Benefits: An Analysis of Incentives in Social Computing. In *Proc. ASIST.*

53. Webster, J. D., McCall, M. E. (1999). Reminiscence functions across adulthood: A replication and extension. *J. Adult Dev.*, 6(1):73–85.

54. Welser, H. T., Cosley, D., Kossinets, G., Lin, A., Dokshin, F., Gay, G., Smith, M. (2008). Finding social roles in Wikipedia. *American Sociological Association*, Boston.

55. Welser, H. T., Cosley, D., Kossinets, G., Lin, A., Dokshin, F., Gay, G., Smith, M. (2011). Finding Social Roles in Wikipedia. *iConference 2011.* (63%, best paper)

56. Yuan, Y. C., Cosley, D., Ling, X., Welser, T. (2009). The Diffusion of a Task Recommendation System to Facilitate Contributions to an Online Community. *International Communication Association.*

57. Yuan, Y. C., Cosley, D., Welser, H. (2007). The Impact of Network Relations on the Diffusion of SuggestBot in Wikipedia. *National Communication Association*, Chicago, IL.

58. Zhao, O. J., Ng, T., Cosley, D. (2012). No forests without trees: particulars and patterns in visualizing personal communication. *iConference 2012.* (32%)

59. Zhao, X., Salehi, N., Naranjit, S., Alwaalan, S., Voida, S., Cosley, D. (2013). The Many Faces of Facebook: Experiencing Social Media as Performance, Exhibition, and Personal Archive. *CHI 2013.* (20%, best paper)

60. Zhao, X., Sosik, V. S., Cosley, D. (2012). It's Complicated: How Romantic Partners Use Facebook. *CHI 2012.* (23%)