Crowdsourcing and service delivery

H. Jamjoom
H. Qu
M. J. Buco
M. Hernandez
D. Saha
M. Naghshineh

Today, service delivery organizations operate in a highly dynamic, complex, competitive, and globally distributed environment. There is constant pressure to reduce costs and improve performance and quality. Success demands the ability to continuously learn and adapt. Standardization is recognized as essential to reducing variation and, therefore, costs, as well as managing quality. Public frameworks and standards, such as ITIL® (Information Technology Infrastructure Library), provide best-practice guidance and a common vocabulary for managing IT (information technology) services. In this paper, we explore the role that social networking and "crowdsourcing" can play in socializing and developing best practices for a service delivery organization. We draw on our experience developing and deploying a social networking application, called ***Cyano****, which is being used by approximately 13,000 IT professionals to capture and maintain day-to-day activities, processes, and artifacts used for problem and change management of several hundred outsourced infrastructures. Cyano is a new breed of social networking enterprise applications, in which crowdsourcing is leveraged to enrich and maintain IT processes and social networks are not created by explicit membership, but rather are implicitly discovered by the type of activities and infrastructure elements that various users support. In this paper, we focus on 1) the architecture of Cyano for supporting social tagging and linkage across different layers of management applications, 2) process customization and governance, and 3) an automated recommendation system that has been well received by thousands of IT professionals. We also highlight research challenges in this space.*

1. Introduction

In order to be competitive, service delivery organizations must offer an increasingly complex and diverse set of services with consistent quality, reliability, and security. From the customer's perspective, regardless of the location from which the services are delivered and irrespective of whether sourced internally or outsourced, the cost and quality of the services must align with the customer's business objectives. In addition, many, if not most, IT customers require that their service providers conform to one or more of the industry standards. Standards such as ITIL** (Information Technology Infrastructure Library), COBIT, CMMI, ISO/IEC 20000, and eTOM [1] offer best-practice guidance to the service provider and are recognized as key to effective and efficient service delivery. In this paper, we focus on ITIL, which is generally regarded as the de facto standard for IT service management (ITSM).

The need for standardization goes well beyond the level of detail covered by industry standards. In fact, most service delivery teams operate at a level of detail well below what it addressed by ITIL. For example, ITIL describes the process for handling requests for change in general. However, most members of a service delivery team are more likely concerned with how a specific change should be accomplished. For example, a system administrator is likely to be concerned with a best practice for provisioning a new UNIX** server or how to perform a health check of a DB2* server. Standardization

0018-8646/09/$5.00 © 2009 IBM

at this level improves consistency and quality of service and reduces training costs.

Conventional wisdom suggests that best practices should be defined by a team of expert process architects and then deployed across the service delivery organization. This method is well suited for high-level processes, such as those defined by ITIL. However, the shear volume of processes and the diversity of expertise required render this an intractable problem for process architects, who are not involved in the actual delivery of services. Even more difficult is the task of keeping current and improving those best practices over time in a very dynamic environment. A practical solution requires the involvement of those individuals performing the service delivery tasks.

For this reason, we have turned to crowd computing technologies, which are enabling businesses to leverage the collective intelligence of groups to help solve business problems. Such groups can be employees, customers, or even the masses on the Internet. In his book titled "The Wisdom of Crowds," James Surowiecki asserts that aggregating information from groups yields better results than would be obtained by relying on individual members of the groups. He lists the four elements for forming a "wise" crowd as diversity, independence, and decentralization of team members, in addition to the ability to aggregate input into collective decisions [2].

In this paper, we describe the interplay between crowd computing, or "crowdsourcing," and service delivery. We draw from our experience in implementing and deploying Cyano, a social networking application for capturing and leveraging the knowledge and expertise of the entire service delivery team. Our strategy is to begin by defining high-level processes based on industry and IBM best-practice standards. Then, we rely on the subject matter experts (SMEs) in the service delivery community to annotate and link the process artifacts in their respective areas of expertise. Finally, we encourage the community to rate and suggest improvements to grow and refine the best practices over time.

Cyano currently supports 13,000 crowdsourced IT professionals to capture and maintain day-to-day activities and processes related to problem and change management [3] for several hundred outsourced infrastructures of varying complexities. Cyano serves as a collaborative knowledge-authoring platform and handles massive amounts of user-generated content each day. As the user-generated contents help to enrich the business processes, the linkages between users and the contents also imply potential communities of users (e.g., an implicit user community that is identified by the type of activities and infrastructure elements that the users support). To discover such implicit connections and enable better communication channels among Cyano users, we implemented and deployed a social recommendation system that utilizes "random walk with restart" to explore the linkages between users and the relevant subjects (e.g., processes) and compute the personalized neighborhood for each user. As byproducts, the recommendation system can also recommend subject information that users may find interesting, identify IT experts for specific subjects (by recommending relevant users to subjects), and disambiguate subject information (by recommending similar subjects to subjects). Essentially, we look at the content-generating activities in Cyano as a time-evolving bipartite graph with users and subjects being the two kinds of nodes. Recommendations are based on the computed similarity of any two nodes.

This paper presents two principal contributions. First, we introduce an efficient architecture (which is implemented and deployed in Cyano) for supporting social tagging and linking of process artifacts. Second, we demonstrate an automated recommendation system that makes use of latent connections between captured knowledge and Cyano users to find IT experts.

The paper is organized as follows. Section 2 describes the current model of outsourced IT infrastructures, highlighting key points with social networking applications. In Section 3, we cover some challenges in social networking. In Section 4, we introduce Cyano and provide an overview of its architecture and its use of social tagging in day-to-day IT management. Section 5 presents an effective algorithm to discover latent connections within our enterprise social network and highlights our initial findings covering both the created social network and our recommendation system. We present related work in Section 6. In Section 7, we present concluding remarks highlighting key research challenges.

2. Managing IT service management processes

As mentioned earlier, this paper illustrates how social networking and community knowledge can improve the efficiency and effectiveness of service delivery. In describing the role of social networking in this space, we must also place it within the larger context of ITSM, focusing on key intersects that are likely to have an impact on any community-based approach for ITSM.

As commonly known, the past several years have seen an increased emphasis on standardization (to improve consistency of service delivery) in IT management across all software and operational layers [4]. Common to these efforts is an emphasis on three primary criteria: 1) Infrastructures should be cataloged and tracked in centralized repositories (commonly referred to as *configuration management databases* [CMDBs]); 2) any changes to the environment, regardless of the reasons (e.g., fix problem or add capability), should be based on the authorized knowledge in the CMDB; 3) a set of well-defined (and optimized) processes and workflows should govern how users interact with the infrastructure (e.g., installing a specific operating system) [5].

The processes and workflows (shown schematically in **Figure 1**) used to manage the IT infrastructure are critical to managing the IT environment and, therefore, need to be managed, as do the configuration items that make up the infrastructure. This means that processes should also be stored in a central repository that captures linkage between process artifacts and between process artifacts and other configuration items (organizations, tools, people, etc.). It also means that a change management system is required to maintain the quality and integrity of the process artifacts and linkages. A process CMDB (pCMDB) serves as an online repository for the authorized version of processes used in the environment that is used when assessing the impact of changes to processes and the environment.

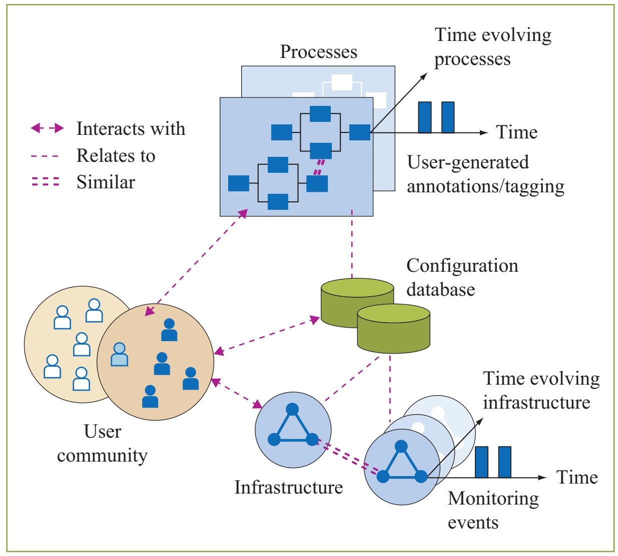


Figure 1

Data model.

3. Challenges in social networking

Reflecting back on Figure 1, there are several key challenges in combining domain knowledge across heterogeneous information sources. These challenges are based on three themes. First, there are very large user communities that are partitioned into many subcommunities, each interacting with different (possibly isolated) portions of the knowledge space. Second, all components are time evolving, making user-generated knowledge ephemeral in nature. Third, as with all community-generated knowledge, there is no global maximum that reflects a global consensus (e.g., what is the best way to apply a patch on Linux**?). However, in enterprise management, having too many alternatives (e.g., forum answers) is a likely distraction. In this paper, we primarily focus on the first challenge. For the other challenges, we rely on the process artifact owner to assess the timeliness of the annotations and select the best practice process based on user ratings and suggested improvements.

Another challenge in social networking is how to encourage participation from the community. According to Jeff Howe, who first coined the term *crowdsourcing*, "The most important component of a successful crowdsourcing effort is a vibrant, committed community [and] understanding what motivates them to contribute" [6]. This is very much a chicken and egg problem. That is, how do you get enough content so that users value it and have the incentive to participate. With Cyano, users were initially required by their respective organizations to document their processes in the repository. We have subsequently added a mechanism to recognize those users who have contributed most to the repository. How to motivate users to contribute is a subject for future research. We believe the key is to establish a community in which participants find value, in which they can have an impact on the state of the practice, and in which they are recognized for their contributions.

4. Cyano process wiki

We have developed Cyano, an enterprise social network application that enables community-based content enrichment. (Cyano is derived from "Cyanoacrylate," the main ingredient in Super Glue**. It is also a shade of blue.) At a high level, Cyano combines a semantic engine, social (semistructured) tagging, and linking of process artifacts with an interactive visualization for published content. It has been deployed globally in IBM and maintains approximately 600 ITSM processes with a user base of approximately 13,000 IT professionals.

Relating the model and challenges described in Sections 2 and 3, our primary goal is to leverage crowdsourcing to enhance and evolve processes artifacts across supported infrastructures. As Cyano users, IT professionals can review the most up-to-date procedures, work instructions, and tools. More importantly, they can also contribute their empirical experience that relates to their job roles as well as rate existing content. **Figure 2** shows a screenshot of Cyano with a work procedure graphically displayed at the upper-right corner and an annotation window opened at the bottom. The user-contributed information and usage patterns are leveraged to provide rich content and recommend related experts to support users' day-to-day activities (further described in Section 5).

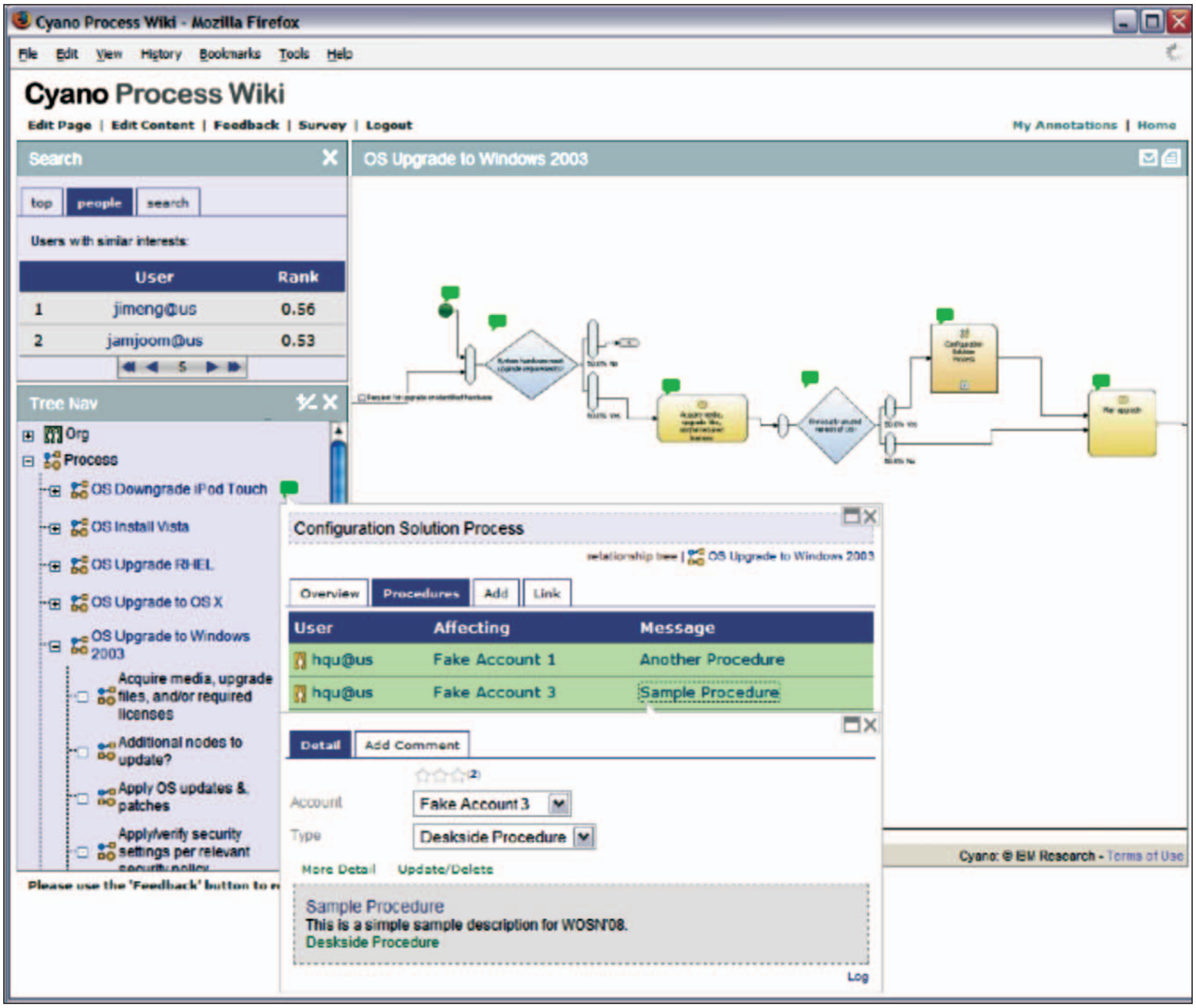


Figure 2

Cyano screen capture.

The initial phase of Cyano development focused on the capture, linkage, and annotation of best practices process artifacts. In subsequent phases, we have been focusing on improving the quality of the process artifacts. To accomplish this, we are adding a governance model and full life-cycle management for process artifacts. Our goal is to decentralize governance to the appropriate level of the organization to avoid bottlenecks and leverage the community expertise. With these enhancements, Cyano will function as a pCMDB for the ITSM processes.

Unlike configuration items for the physical and logical infrastructure, which are usually initialized and audited using automated scans or physical inventories, processes cannot in general be discovered. Discovery of what processes are being used would require that the processes be documented and executed in a workflow engine. This is, of course, the desired end-result, but it is not the state of current practice. Attempts to document process content from top-down generally fall short of capturing the level of detail to be relevant to the service delivery team in the field; thus, it is an ideal candidate for crowdsourcing. Our approach is to have the core of the process infrastructure defined by a team of process architects and then leave it to the subject matter experts in the field to document what they do on a day-to-day basis and link it to the main infrastructure. In order to maintain consistency and quality of the process content, the input is reviewed and rated for completeness and quality and must be approved before it is visible to the entire community.

Crowdsourcing makes possible the creation of content and also facilitates and expedites the evolution of what should be a dynamic set of process artifacts that change and grow to adapt to the changing environment and changing customer requirements. The ability of the community closest to the work to add new content, rate current content, and suggest improvements greatly enhances the vitality of the repository and the community.

Cyano annotations and tagging

In the current deployment, users are organized into different (not necessarily disjoint) groups based on their professional organizations and job responsibilities. Each group typically supports one or more outsourced infrastructures and implements a number of processes for both problem determination and infrastructure change management. As expected from an enterprise application, group membership is strictly controlled and data privacy is maintained across different groups. Currently, there are several thousand groups defined in Cyano.

Depending on the account, technology, and configuration, processes are instantiated to reflect the activities and variations performed on the supported infrastructure. Within a group, users share the same instance of published processes. Only the owners of a process can create a new version of a process. Other users can annotate or tag processes with variations, comments, and recommendations. These annotations are valuable in evolving and maintaining the vitality of the processes. In **Figure 3**, we show an example of how processes are instantiated and tagged by different groups. Process instantiation in Cyano implies maintaining a logical version, where users of the instantiated process can modify and annotate the process using user-defined Web forms. For example, in Figure 3, there are two logical instances being enriched (depicted by the blue and white rectangles). Whenever a user adds or modifies the process, the change is tagged with the logical instance identifier (along with several other attributes). The tagged content, applied to the process template, is then considered a logical instance. Manipulating and maintaining these logical instances to match user interest and privileges in a scalable fashion is the primary challenge in our architecture.

That said, there are four operations (referred to as *annotations*) that are allowed in Cyano: 1) filling a user-created Web form, 2) changing a filled form, 3) adding a task, or 4) disabling a task. Annotations can be associated with any content element (e.g., process task). They are additive, where the *derived* view is a union of all annotations on every content element in the view. This is depicted in Figure 3, where the annotations from the two logical instances are merged in the bottom process.

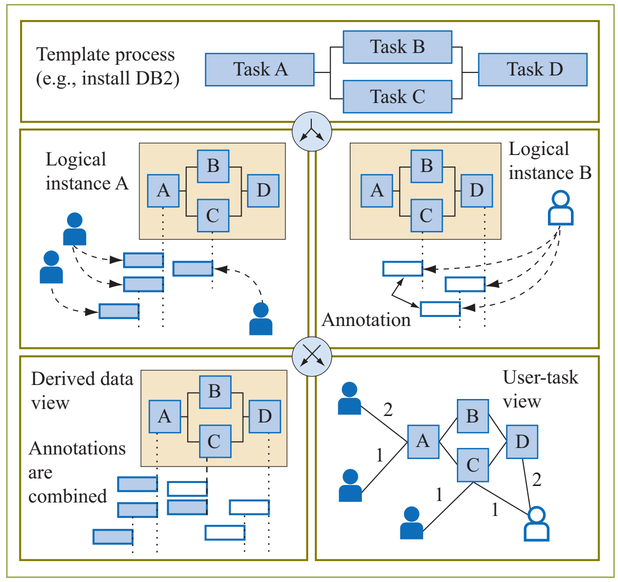


Figure 3

Process tagging in Cyano.

Besides user-generated tags, Cyano makes extensive use of system tagging to maintain its content and track user activities. These system-generated tags describe the context of the annotation. For example, in Figure 3, the logical instance identifier is a tag that is associated with all user-generated annotations. Maintaining logical instances as well as creating derived views is then a union of annotations that have been prefiltered against user tags and system tags.

Annotation or tagging in Cyano enables users to customize (i.e., record variations) to existing processes and additionally to restrict the variation to a particular account or group of accounts or geography. This eliminates the proliferation of many processes with only slight variations and maintains the association between the standard processes and any variations. Facilitation of customization may be contrary to the goal of standardization in service delivery. However, realities of a very large and diverse customer set dictate that a certain level of customization is required. The goal is to limit customization as much as possible and to document and link process variations to the authorized standard. In this way, customization can be managed and the set of standardized processes can be minimized. What is to be avoided is a large set of unrelated processes with no relationship to the established standards.

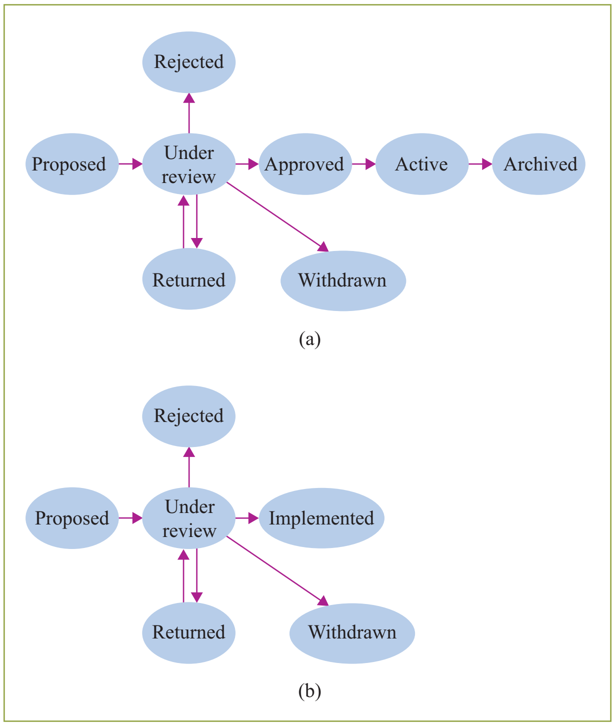


Figure 4

Life cycles: (a) process artifact life cycle; (b) RFC (Request for Change) life cycle.

Cyano linkages

One of the most important features of a pCMDB is the linkages it contains between configuration items. The network of typed and directional links is what enables impact assessment, which is an essential component of change, problem, and incident management. In Cyano, linkages between process artifacts and between process artifacts and other items (e.g., roles, tools, standards, organizations, and geographies) enable an impact assessment of changing a process artifact. In Cyano, users can add links, for example, link a process artifact to a particular technology, product, organization, or account. These links are reviewed and approved by the owner of the source of the link prior to being made visible to the community in order to maintain quality and consistency of the repository.

In addition to creating linkages between process artifacts and between process artifacts and other items, Cyano users are encouraged to customize their experience in Cyano by creating user profiles that specify their interests. The user profile is essentially a set of links between the user and process artifacts or other items. For example, a user may specify exactly which process artifacts are of interest, on what accounts he works, and which technologies are of interest. This information is used to customize the views presented to the user, to assess who will be affected by changes to process artifacts, and to notify users of changes or additions to the repository that are likely of interest.

Cyano life-cycle management

As with all configuration items in a CMDB, process artifacts in Cyano need to be managed over their full life cycle. **Figure 4(a)** depicts a possible life-cycle state diagram for a process artifact. For process artifacts in certain protected states, such as "Approved" and "Active," an RFC (Request for Change) is required to make any change to a process artifact. A change management process is needed to review each RFC and determine the potential impact prior to approving and subsequently implementing the requested change. **Figure 4(b)** shows a possible life cycle for RFCs. When significant changes are made to a process (i.e., new or deleted tasks, change in the process flow), it is recommended that a new version of the process artifact be created. For any change, interested parties should be notified. Additionally, it is helpful to request feedback and track the implementation of changes in the environment.

Cyano implementation

Cyano was fully implemented using a combination of PHP (www.php.net), AJAX (AJAX combines Asynchronous JavaScript** and XML, and it is commonly used to drive interactive user experience), and MySQL database. Cyano also interacts with diverse external data sources including process models authored through a variety of tools (e.g., Visio**, Dia, and WebSphere Business Modeler*). The Cyano architecture is built on top of a service data object (SDO) layer, called *managed objects* (MO). The key difference between traditional SDO implementations, which abstract out databases into persisted data objects [7], and MO is that MO allows for user-defined semantic linkage. (While the current deployment of Cyano does not include infrastructure linkage, adding this capability is a logistical issue, rather than a design limitation.) MO gives users the flexibility to create their own linkage across different knowledge components, which is essential in an enterprise environment with diverse data sources. The separation of link management and the original data sources also helps support various levels of access control across outsourced infrastructures.

Leveraging crowdsourced process knowledge

As with any social network, knowledge collection is only the first step. More opportunities arise in utilizing the collected information and maintaining healthy growth of the community. In our application, we focus on the following key challenges:

Expert recommendation—We explore how mining historical behavior of users can identify experts on specific topics. Here, we assume that authorship and constant annotations on a task indicate users' interests; usage of annotations reflects the expertise of their authors. We note that expert recommendation encompasses identifying users of similar interests. This mirrors the goals of mainstream social networks.

Disambiguation—Grouping similar tasks is important in task disambiguation and information search. With users adding their knowledge independently into Cyano, there is a great possibility of redundant information. Given that a fully automated merge functionality is nearly impossible, task grouping acts as a crucial initialization step before merging. Additionally, during search, task grouping helps support "fuzzy" searches, which is well beyond what is typically provided by exact word matching. For example, additional tasks similar to those in the word-matching results set could be returned to enhance the search function.

The above challenges can be addressed by designing an accurate and intuitive similarity measure among users and tasks based on users' annotation and content browsing behaviors. Because annotations usually represent a stronger linkage than browsing, we progress from analyzing user annotations toward measuring the similarity among users and contents. We construct an undirected graph based on user-task annotations where the edge weight is computed based on annotation frequency. This is illustrated in **Figure 5**.

In Cyano, we adopt the *random walk with restart* mechanism [8, 9], which utilizes network connectivity between users (content creators) and tasks (hosted content) to find expert users on specific tasks, to identify similar tasks, and to compute the personalized neighborhood for each user. This algorithm provides a simple and elegant solution that automatically leverages the latent connections of different processes for social recommendation. Similar approaches have been successfully applied to many different applications which are described in Section 6.

Intuitively, we repeat random walks starting from a query node (either a user or a task). Suppose a random particle sets out from a query node and starts jumping around on the graph. The probability of the particle jumping to the adjacent nodes is proportional to the edge weights. There is also a probability that the particle jumps back to the query node, the so-called *restart*. The number of times that each node on the graph is visited reflects the relevance of those nodes to the query node *i*. Specifically,

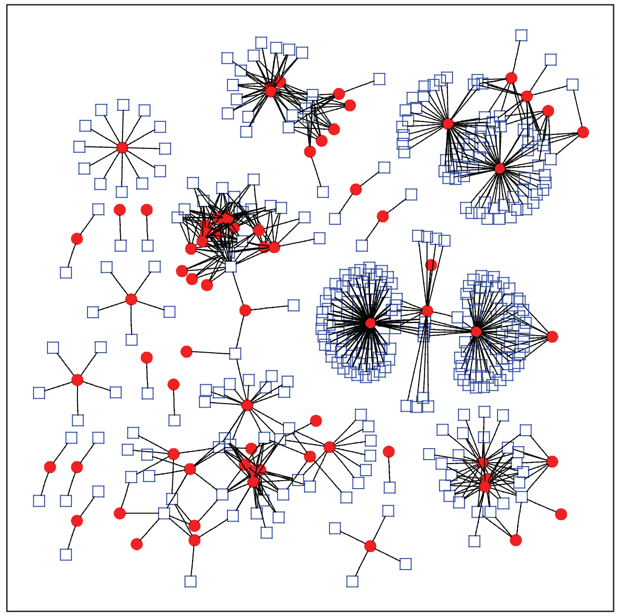


Figure 5

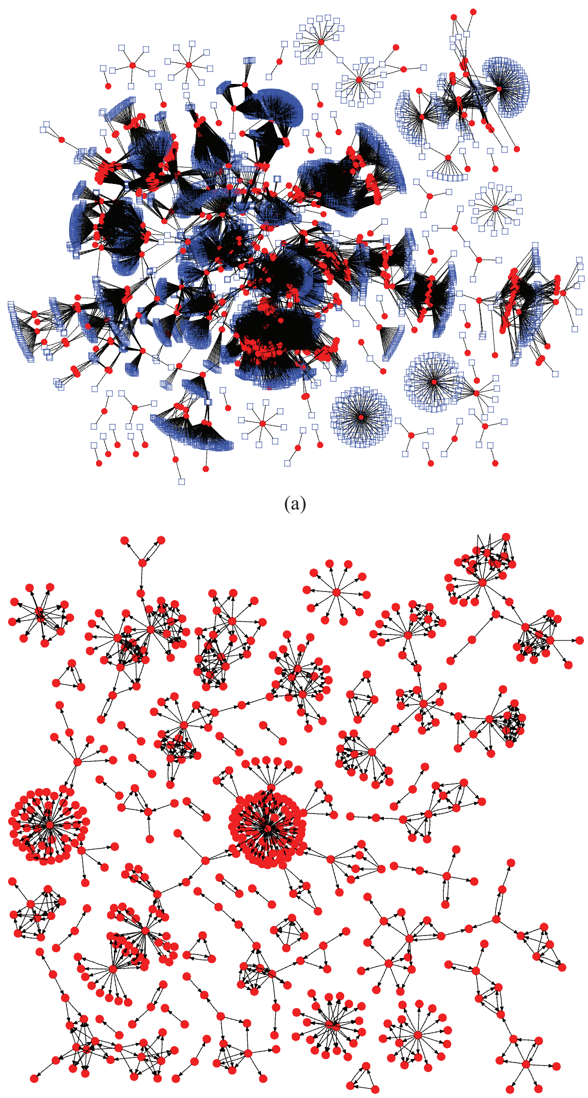
User-task-editing networks (January 2008): user (circle), task (square), edge weight (annotation frequency).

the relevance score between the query node *i* and node *j* is the steady-state probability that the random particle will remain at *j*. This algorithm naturally incorporates connections through multiple nodes yet favors shorter connections and lower-degree nodes.

Visualizing the results

In this section, we visualize the social network data we collected through Cyano, over a five-month period. The graph visualization uses the Kamada-Kawai algorithm [10] and is generated by Graphviz** [11]. We also present initial results from the recommendation algorithm.

To illustrate the graph formation, we first showed in Figure 5 a small dataset derived from user-to-task annotations in January 2008. This dataset is small in nature because January usually carries light editing traffic due to the extended holiday. The dataset has 1,438 annotations that are contributed by 71 users on 349 tasks. To improve figure readability, edge weights, which reflect annotation frequency, are hidden. Task-task relationships (relevance among tasks) are also omitted since we are concentrating on the analysis of user-task interactions. The user nodes are denoted by solid circles and task nodes by empty boxes. Three primary observations can be seen in this figure: 1) Many users annotate the same tasks as other users, and some even form a group due to shared interest; 2) there are users who have no overlap



(a)

(b)

Figure 6

Social behavior and user neighborhood network: (a) user-task-editing network (September 2007–December 2007): user (circle), task (square), edge weight (annotation frequency); (b) user-user neighborhood network (top 1,000 weighted edges): user (circle), edge weight (computed relevance score).

with others even though they actively annotate many tasks; 3) although there are several high-degree user nodes (i.e., very active or omniscient users), there are no tasks that are annotated by a high number of users. This is different from many other social media such as Wikipedia** [12] and Flickr** [13], where skewed distributions exist for both users and contents.

Table 1 Comparison of several similarity measures.

Method	*Number of connections*
Random-walk-restart	3.88e+5
Jaccard	0.67e+5
Cosine	0.67e+5

Nonetheless, this may change if we increase the granularity of the contents, such as analyzing user-to-process annotations (process is a sequence of tasks).

Figure 6 presents the social network and its analysis on user neighborhood formation during September 2007 to December 2007. Figure 6(a) shows the user-task annotation network, which includes 44,416 annotations that are contributed by 792 users on 4,999 tasks. The figure clearly shows many hotspots near the center area, where many users share similar interests. It also confirms the three observations highlighted in Figure 5.

Figure 6(b) shows the user neighborhood network based on the same dataset described in Figure 6(a). This is shown as a weighted directed graph. User A pointing to User B with edge weight 0.9 means that User A is similar to User B with a relevance score of 0.9. To increase graph readability, we hide edge weights and only plot edges belonging to the top 1,000 relevant scores. Figure 6(b) demonstrates two observations: 1) There are *star users* who share similar interests with many other users, such as the user in the center of the figure with edges pointing to hundreds of other users; 2) the neighborhood relationship is not necessarily symmetric, especially near star users. This is because the relevance score computed by random walk with restart is not symmetric. The steady-state probability to reach a node with a high degree is usually higher than to reach a node with a low degree.

We have also compared our algorithm against other similarity measures. Our preliminary results are shown in **Table 1**. Note that random walk with restart is able to identify 5 times more connections than traditional methods, including Jaccard coefficient and cosine similarity. This confirms the effectiveness of the algorithm in revealing hidden connections, which is crucial in our sparsely connected enterprise social networks. A more detailed evaluation is the topic of upcoming work.

Related work

Social networks have recently attracted significant research attention because of their extreme popularity on the Internet, where one of the most successful functions is social tagging. Studies [14] have focused on how to recommend the most appropriate tags to users. Unlike short or simple tags on the Internet, tagging in enterprise applications often requires more complex annotations. With data from large social networks becoming readily available, there is also an increased focus on building models describing various aspects of social networks, such as network evolution [15], community characteristics [16], and visualization [17].

Our work focuses on a different aspect, namely how to leverage relationships between the workforce and the supported infrastructure in order to improve content capture and expert recommendation. One closely related area is the proximity measure on social networks presented as graphs. We use random walk with restart to model the similarity between two nodes. This method is related to page rank [18] and its variants. Random walk with restart has been used widely in many applications: Mixed Media Graph (MMG) on bio-image retrieval [19], SimRank on Web graphs [20], neighborhood formation on bipartite graphs [21], and center-piece subgraph for graph clustering [22].

Our work also relates to enterprise process management, where it is dominated by widely used commercial systems including (IBM MQSeries*, SAP, Oracle**, etc.). Many research efforts have focused on framework design and process integration, transformation, and verification. Examples of such research include process transformation and integration [23] and verification framework for Web services [24]. Closely relating to our work, Yang et al. [25] propose a social network framework for Web 2.0 applications. Our work has a different focus on leveraging social network and data mining for outsourced IT infrastructures.

Conclusions and challenges

In this paper, we presented our experience in leveraging crowdsourcing in capturing and enhancing service delivery process content. Not surprisingly, there are still several open issues in the application of social networking in IT management. As we highlighted, social networking and crowdsourcing can have a tremendous impact on information discovery and enrichment. Given the large amount and time-evolving nature of related content, tracking the implication of information variability is an open challenge. There are also fundamental challenges in leveraging the relationships between practitioners and the underlying infrastructures they support. For example, if two practitioners provide database support, it is still unclear when they should be considered part of the same social community, given that databases (even from the same vendor) may have large variations in configurability and maintainability. This is further complicated by the security and data privacy implications when crossing organizational boundaries (e.g., in two separate outsourced infrastructures). Finally, there are also fundamental challenges in creating robust models to motivate community participation for process enrichment.

*Trademark, service mark, or registered trademark of International Business Machines Corporation in the United States, other countries, or both.

**Trademark, service mark, or registered trademark of the U.K. Office of Government Commerce, The Open Group, Linus Torvalds, Henkel Corporation, Sun Microsystems, Microsoft Corporation, AT&T Corporation, Wikipedia Foundation, Inc., Yahoo!, Inc., or Oracle Corporation in the United States, other countries, or both.

References

1. *ITIL Design Guidelines: Service Operation Version 3,* TSO for OGC, 2007; see *http://www.itilversion3.com/*.
2. J. Surowiecki, *The Wisdom of Crowds,* Anchor Books, New York, 2005.
3. *ITIL: The Key to Managing IT Services: Service Support Version 2.3*, TSO for OGC, 2000; see *http://wareseeker.com/Business-Finance/itil-toolkits-2.3.zip/12899*.
4. M. B. Kelly, "The TeleManagement Forum's Enhanced Telecom Operations Map," *J. Network Syst. Management* **11**, No. 1, 109–119 (2003).
5. L. Klosterboer, *Implementing ITIL Configuration Management*, IBM Press, 2008.
6. J. Howe, *Crowdsourcing Why the Power of the Crowd is Driving the Future of Business,* Crown Business, New York, 2008.
7. BEA Systems, IBM, Oracle, Primeton Technologies Ltd., Rogue Wave Software, SAP AG, Software AG, Sun Microsystems, Sybase, Xcalia, Zend Technologies; *Service Data Objects*; see *http://www.ibm.com/developerworks/library/specification/ws-sdo/*.
8. J. Sun, H. Qu, D. Chakrabarti, and C. Faloutsos, "Relevance Search and Anomaly Detection in Bipartite Graphs," *SIGKDD Explorations,* Special Issue on Link Analysis, pp. 48–55, 2005.
9. J. Sun, H. Qu, D. Chakrabarti, and C. Faloutsos, "Neighborhood Formation and Anomaly Detection in Bipartite Graphs," *Proceedings of the International Conference on Data Mining (ICDM'05)*, pp. 418–425, 2005.
10. T. Kamada and S. Kawai, "An Algorithm For Drawing General Undirected Graphs," *Information Processing Lett*. **31**, No. 1, 7–15 (1989).
11. Graphviz: Graph Visualization Software; see *http://www.graphviz.org/*.
12. F. B. Viegas, M. Wattenberg, J. Kriss, and F. van Ham, "Talk Before You Type: Coordination in Wikipedia," *Proceedings of the 40th Annual Hawaii International Conference on System Science*, p. 78, 2007; see *http://www.research.ibm.com/visual/papers/wikipedia_coordination_final.pdf*.
13. T. Rattenbury, N. Good, and M. Naaman, "Towards Automatic Extraction of Event and Place Semantics from Flickr Tags," *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 103–110, 2007.
14. Z. Xu, Y. Fu, J. Mao, and D. Su, "Towards the Semantic Web: Collaborative Tag Suggestions," *Proceedings of the Collaborative Web Tagging Workshop at the World Wide Web Conference*, 2006; see *http://www.semanticmetadata.net/hosted/taggingws-www2006-files/13.pdf*.
15. J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graphs Over Time: Densification Laws, Shrinking Diameters and Possible Explanations," *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge Discovery in Data Mining*, pp. 177–187, 2005.

16. R. Kumar, J. Novak, and A. Tomkins, "Structure and Evolution of Online Social Networks," *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 611–617, 2006.
17. M. Dubinko, R. Kumar, J. Magnani, J. Novak, P. Raghavan, and A. Tomkins, "Visualizing Tags Over Time," *Proceedings of the 15th International Conference on World Wide Web*, Edinburgh, Scotland, pp. 193–202, 2006.
18. S. Brin and L. Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine," *Computer Network ISDN System* **30**, No. 1-7, 107–117 (1998).
19. J. Pan, H. Yang, P. Duygulu, and C. Faloutsos, "Automatic Multimedia Cross-Modal Correlation Discovery," *Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Seattle, WA, 2004, pp. 653–658.
20. G. Jeh and J. Widom, "SimRank: A Measure of Structural-Context Similarity," *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 538–543, 2002.
21. H. Qu, J. Sun, and H. Jamjoom, "SCOOP: Automated Social Recommendation in Enterprise Process Management," *Proceedings of IEEE International Conference on Services Computing (SCC'08)*, pp. 101–108, 2008.
22. H. Tong and C. Faloutsos, "Center-piece Subgraphs: Problem Definition and Fast Solutions," *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 404–413, 2006.
23. K. Fujiwara, B. Ranachandran, A. Koideand, and J. Benayon, "Business Analysts and Business Process Transformation Technology," *IEEE International Conference on Services Computing*, pp. 83–90, 2007.
24. S. Moser, A. Martens, K. Gorlach, W. Amme, and A. Godlinski, "Advanced Verification of Distributed WS-BPEL Business Processes Incorporating CSSA-Based Data Flow Analysis," *IEEE International Conference on Services Computing*, pp. 98–105, 2007.
25. S. J. H. Yang, J. Zhang, and I. Y. L. Chen, "Web 2.0 Services for Identifying Communities of Practice through Social Networks," *IEEE International Conference on Services Computing*, pp. 130–137, 2007.

Received April 20, 2009; accepted for publication June 30, 2009

Hani Jamjoom *IBM Research Division, Thomas J. Watson Research Center, 19 Skyline Drive, Hawthorne, New York 10532 (jamjoom@us.ibm.com).* Dr. Jamjoom is a Research Manager at IBM T. J. Watson Research Center, where he is focusing on the operational side of service science management and engineering (SSME), covering topics in data mining, social networking and collaboration, knowledge management, and system scalability. Dr. Jamjoom received a B.S. degree from the Rose-Hulman Institute of Technology, Terre Haute, Indiana, in computer engineering and an M.Eng. degree from Cornell University, Ithaca, New York, in electrical engineering. He received the Ph.D. degree in computer science in 2004 from the University of Michigan, Ann Arbor, where he focused on quality-of-service architectures and the integration of controls in networks and operating systems to manage Internet services during overload scenarios.

Huiming Qu *IBM Research Division, Thomas J. Watson Research Center, 19 Skyline Drive, Hawthorne, New York 10532 (hqu@us.ibm.com).* Dr. Qu is currently a postdoctoral scientist in services research at the IBM T. J. Watson Research Center. She received her Ph.D. degree in computer science from the University of Pittsburgh in 2007. Her research has been focused on Web databases, real-time databases, mobile databases, and graph mining.

Melissa J. Buco *IBM Research Division, Thomas J. Watson Research Center, 19 Skyline Drive, Hawthorne, New York 10532 (mjbuco@us.ibm.com).* Ms. Buco is a Senior Software Engineer in the Service Delivery department at the IBM T. J. Watson Research Center. She received a B.A. degree in mathematics from Northeastern University and an M.S. degree in computer science from Columbia University. Her recent focus has been in the area of IT service management. She has also worked in the areas of service-level agreement management, software engineering, emergency management, project management, and workflow management.

Milton Hernandez *IBM Research Division, Thomas J. Watson Research Center, 19 Skyline Drive, Hawthorne, New York 10532 (miltonh@us.ibm.com).* Mr. Hernandez is an innovator with 19 years of hands-on experience in the IT industry and 4 years in services enablement research. He is currently a Senior Technical Staff Member in the IBM Research Services Enablement organization. He leads a number of global initiatives in identity and access management, enhancing customer service experience and quality. Prior to his current role, he spent more than 15 years in IBM Global Services developing solutions and supporting customers applying various technologies across multiple industries.

Debanjan Saha *IBM Systems and Technology Group, 9000 South Rita Road, Tucson, Arizona 85744 (dsaha@us.ibm.com).* Dr. Saha is an innovator and an entrepreneur with more than 15 years of experience in the IT industry. He is currently Director of Storage Software and Solution Enablement in the IBM Systems and Technology Group. In this role, he leads a global team spanning multiple sites and is responsible for development execution of various storage products including SVC (SAN Volume Controller), SoNAS (Scale-out NAS), Hydra, and DS8000* advanced functions. Prior to his current role, he spent several years in IBM Research and in Tellium, Inc., an optical networking pioneer, which he helped grow from an early stage startup to a public company. He is a Fellow of the IEEE (Institute of Electrical and Electronics Engineers), a Distinguished Scientist of the ACM (Association of Computing Machinery), and an IBM Master Inventor. He holds a Ph.D. degree in computer science from the University of Maryland at College Park.

Mahmoud Naghshineh *IBM Research Division, Thomas J. Watson Research Center, 19 Skyline Drive, Hawthorne, New York 10532 (mahmoud@us.ibm.com).* Dr. Naghshineh is Director of Services Delivery Research, with worldwide responsibility for next-generation IT services technologies, processes, and tools. He leads global research efforts in support of strategic outsourcing, managed business process services, IT optimization, and transformation business. From 2004 to 2005, he was a partner at IBM Business Consulting Services, responsible for the federal industry alliances account. From 2002 to 2004, he was the director of emerging markets at the CTO office, IBM Systems and Technology Group. From 1990 to 2002, he worked at the IBM Research Division in software and services related to Web-based infrastructure, mobile and wireless, secure computing platforms, telecommunications services, and quality-of-service provisioning. He joined the IBM Systems Group in 1988. In 2002, he was elected an IEEE Fellow. He has led industry groups and NSF and government research conferences and workshops. He has served as the editor-in-chief of IEEE Wireless Communications Magazine, program co-chair of MobiCom 2001, and chairperson for many IEEE/ACM events. Dr. Naghshineh was an adjunct professor at the department of electrical engineering, Columbia University, from 1997 to 2001. He has published more than 100 technical papers and holds a number of IBM Outstanding Recognition Awards and patents. He received his doctoral degree form Columbia University.