



<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper presented at *36th Information Systems Research Seminar in Scandinavia*.

Citation for the original published paper:

Wingkvist, A., Ericsson, M. (2013)

Asked and Answered: Communication Patterns of Experts on an Online Forum.

In: *Proceedings IRIS36: August 11-14 2013 at Gran, Norway*

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:lnu:diva-33541>

Asked and Answered: Communication Patterns of Experts on an Online Forum

Anna Wingkvist¹ and Morgan Ericsson²

¹ Linnaeus University, Växjö, Sweden
`anna.wingkvist@lnu.se`

² University of Gothenburg/Chalmers, Gothenburg, Sweden
`morgan.ericsson@chalmers.se`

Abstract. The notion of network structure for social relations dates back half a century. Nowadays people form social networks offline as well as online. At an online community people are connected through information exchange of sorts. Interest groups form often forums to aid each other and discuss things. Programmers are no exception and a question and answer site called Stack Overflow has been up and running since 2008. Our focus is to find patterns of how people interact on this online community and see if we can find expert users. We find 4 different ways to categorize experts, which result in different rankings. We also investigate how expertise is divided among topics, and find some overlap with the global ranking.

Key words: social network, online community, social network analysis, help seeking, programming, communication patterns, expert finding

1 Introduction

People use Internet technologies to communicate with each other. When enough people share a common interest sometimes online communities are formed. An online community can be regarded as a social network. Boyd and Ellison [3] define a social network site as a web-based service that allow people to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users, and (3) view their list of connection. To study these online communities, researchers apply social network analysis [9], often supported by visualizations to illustrate the findings [7]. Social networks analysis focuses on discovering patterns of relationships between people (actors), who are part of the social network. Connections between actors are called ties. This approach also focus on how information flows between actors (e.g., what kind of information, between whom, to what extent, etc.). This is done to identify informal networks and quantify through statistical analysis the results. Robins [11] suggests that a social network can be represented and visualized as a graph.

Online communities have emerged as one of the most important places for people to seek advice or help [13]. Users of these social network sites usually do not know each other and often are only identified by alias. Still they

are willing to interact with each other for various reasons, such as altruism, reputation privileges, expected reciprocity and direct learning benefits according to Kollock [10]. Common interests are for example based on health, food, sport, programming etc. One example of an online community is Stack Overflow (<http://www.stackoverflow.com>) — it is a forum where (software) developers from different backgrounds can exchange information in the form of questions and answers.

Users of forums have different levels of expertise, and this can affect the quality of their answers. For example, an expert user should be more likely to provide a correct answer. However, since users do not know each other, it can be difficult to determine who is an expert. In this paper we investigate if an analysis of communication patterns can be used to determine expertise. We discuss and investigate different possible patterns and compare which users are considered experts by these patterns. We acknowledge that users can be experts in various fields, and compare how the different patterns rank users for all topics as well as for selected topics. All analyses are performed on a data set from Stack Overflow.

2 A study of the Stack Overflow structure

In order to understand how developers behave and what they talk about on online forums, we need data to study. We rely on a data set, from Bacchelli [1], which contains all the activity on Stack Overflow between 2008-07-31 and 2012-07-31. The data set contains approximately 1.3 million users, 3.5 million questions, 7 million answers, and 13 million comments. Each of the questions are tagged with one or more of approximately 30,000 topics.

2.1 Users and Usage

There is a significant difference between users; if we only consider active users, about 47% have only asked questions and 25% have only answered questions. The bowtie structure [5] is often used to describe complex network structures. It consists of four distinct parts: *In*, *Out*, *Core*, and *Tendrils and Tubes*. *Core* forms the strongly connected center of the network structure where every node is somehow connected. *In* and *Out* links only to and from the core, respectively. *Tendrils and Tubes* consists of nodes that is not part of the *Core*, but links to or from the *In* or *Out*. If we use the bowtie structure to describe the forums *Core* would be the users involved in active discussion, *In* the users that have their questions answered by the *Core*, and *Out* the users that answer questions asked by the *Core*. The *Tendrils and Tubes* are users that do not interact with the *Core*.

The bowtie structure of the world wide web consists of four approximately equal parts, while Stack overflow has a much larger *In* than *Core* and *Out*, respectively (c.f. Figure 1). Zhang et al. [13] report similar results in their study of the Java Forum network and suggest that a smaller *Core* part can be interpreted

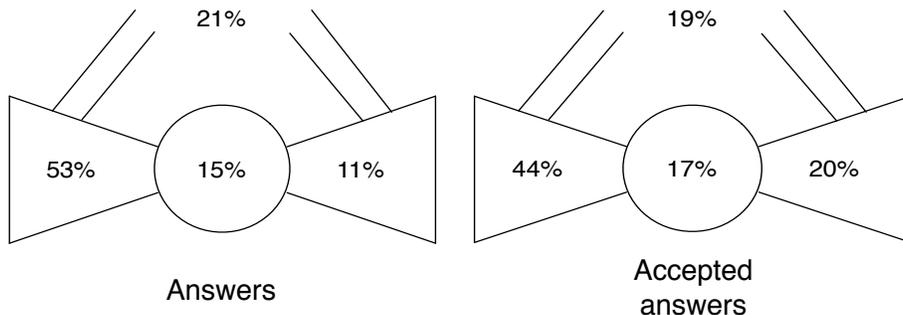


Fig. 1. The structure of the accepted answers (left) and answers (right) networks represented as bowties. The percentages indicate the number of users that are in the In, Core, Out, and Tendrils and Tubes.

as that a forum is not a community where people discuss and help each other, but rather a public place to get questions answered by volunteers.

If we consider user activity based on how many questions they ask and/or answer, it is also varies a lot. A majority of the users have never answered a question, and only 15% and 25% have answered or asked five or more questions. However, there is a group of users that have asked and/or answered more than a thousand questions. The network formed by questions and answers is a scale-free network [2] (the degree distribution follows a power law [6]), i.e., there exist preference — questions and answers are not asked or answered on random.

There is no bias with respect to asked/answer ratio when users decide which questions to answer. Users with a high degree of answered questions answer questions from users with no to many answered questions. This may seems contradictory to a network with preference, but this is not the case. It simply follows from the fact that most users ask and answer few questions and a majority of the questions and answers are from relatively few users.

A user that asked a question can decided which answer was the best or most correct by *accepting* the answer. If we only consider questions and accepted answers, we find a bowtie structure that is similar to the previous one (c.f. Figure 1). Since there is at most one accepted answer per question, the number of users that have only asked questions has increased, while the users that have only answered questions has decreased. The Core discussion and Tendrils and Tubes are more or less unchanged. The network formed by questions and accepted answers is a scale-free network as well.

There also seems to be no bias with respect to the number of answered when users decide which answer to accept. If we consider the accepted/answer ratio, we find a positive skew; values between 0.0 and 0.5 are more common than values between 0.5 and 1.0. It is interesting to note that there exist users with more than 500 answers and a accepted/answer ratio above 0.8.

Users can vote on posts, for example voting it up or down, and this results in a score per post. This score is an integer, which ranges from -132 to 4432 in

the data set. If we consider the total score and number of posts per users, we find a linear relationship between the two. There is nothing that suggests that a user with a higher amount of posts would have a higher per-post score, or the opposite.

2.2 Tags and topics

Each question can be tagged with up to 5 topics. It is up to the user asking the question to either pick among the existing or introduce new topics. The topics range from generic ones, such as *database* or *algorithm* to quite specific, such as *mysql-error-1067* or *jquery-1.7*.

The tags and topics are not structured like a bowtie, but rather like a Jellyfish [12]. There is a complete subgraph that forms the center of the topic network, and topics that are not part of the central core form rings that link to either the central core or a ring. Given the size of the topics network (more than 30,000 nodes and 1.75 million edges), it is computationally expensive to determine the size of the complete subgraph and the rings.

There are no obvious patterns in the topic graph. Given the large amount of edges, many topics are connected to 25-50% of the total topics. We applied a community finding algorithm (Edge-betweenness [8]) and could identify 4 distinct topic groups. These are discussed during the experiment in Section 3.3.

If we consider the topics that are connected to the most other topics, we for example find Java, c# and Android. If we instead consider patterns over time, i.e., the topics that receive the most new questions and answers every month, we also find Java, c# and Android.

3 Finding Experts

Given the large amount of available data from Stack Overflow, can we find experts by looking at communication patterns? It is not clear which communication patterns we expect an expert to have, so we begin by discussing various such patterns and we then search for these patterns in the data to rank users. We then compare these rankings to see if different patterns produce similar rankings, which can be interpreted as an agreement.

3.1 Defining an Expert

The large amount of data makes it impractical to define an expert based on the content of the posts, so we rely on the network structure and specific attributes, such as up-votes or accepted answers. With these constraints, a simple definition of an expert could be the number of questions or users the person has answered. The higher this number, the more of an expert the user is.

The number of answers can be an imprecise measure, since we do not know how relevant the answers are. There is a moderation system on Stack Overflow,

where posts can be marked as spam or irrelevant, and since we filter any such posts, we assume that all answers are relevant. We can also apply some of the post attributes to help improve our measure of how relevant an answer is, for example by only considering accepted answers or answers with a high score.

Accepted answers can be misleading, since there can only be one accepted answer per question, so if two or more users post answers, only one of these will be accepted. We cannot assume that the best answer is always the accepted one; it could simply be that the first answer that is good enough is accepted. High total scores can also be misleading, since the total score will grow with the number of posts, so a user with a large amount of posts where each has a low score, will still have a high total score. We can apply the average score, or the ratio of accepted compared to answered. However, this again can be misleading, depending on how we define an expert. Is it better to have many posts, even if few of them have high scores or are accepted, than only have a few posts but with high scores or accepted?

To summarize, we consider the following as possible attributes of an expert:

- A large amount of answers
- A large amount of accepted answers
- A high total score
- A high average score
- A high ratio of accepted against answered

When we discuss different attributes above, we do not take questions posted into account or relations to other experts. Is a person that posts equally many questions and answers still an expert, or should there be a larger amount of answers? Is a person that answers questions of other experts more of an expert? We can add the following attributes to the list.

- A majority of answers
- Should answer other experts

3.2 Searching for Experts

Based on the attributes defined in the previous section, we can not define a set of actual measures of an expert and compare these. For each measure we find the top 50 experts. These sets of experts are compared among the different measures, in order to find how well they correlate. Note that there is no actual list of expertness, so we define correlation as a measure of correctness; the more methods that correlate, the more correct the results are.

We define the following analyses on the network:

1. Extract the top 50 users based on how many other users they have replied to.
2. Extract the top 50 users based on how many answers they have written in total.
3. Extract the top 50 users based on accepted answers to users.

	1	2	3	4	5	6	7	8	9	10
1	1.0 (50)	0.38 (47)	0.26 (39)	0.22 (38)	0.16 (37)	0.0 (0)	0.0 (0)	0.0 (0)	0.18 (37)	0.22 (37)
2		1.0 (50)	0.28 (40)	0.3 (40)	0.26 (37)	0.0 (0)	0.0 (0)	0.0 (0)	0.2 (38)	0.24 (38)
3			1.0 (50)	0.44 (45)	0.16 (34)	0.0 (0)	0.0 (0)	0.0 (0)	0.16 (30)	0.16 (33)
4				1.0 (50)	0.18 (35)	0.0 (0)	0.0 (0)	0.0 (0)	0.18 (31)	0.16 (34)
5					1.0 (50)	0.0 (0)	0.0 (0)	0.0 (0)	0.2 (41)	0.18 (32)
6						1.0 (50)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)
7							1.0 (50)	0.0 (0)	0.0 (0)	0.0 (0)
8								1.0 (50)	0.0 (0)	0.0 (0)
9									1.0 (50)	0.16 (36)
10										1.0 (50)

Table 1. The results of the 10 analyses compared. The first figure shows how similar to sequences are by computing dividing the number of matches with the total number of elements (from 0.0 to 1.0) and the second number shows the number of elements that were in both sequences (from 0 to 50). Note that the table is mirrored across the middle.

4. Extract the top 50 users based on accepted answers they have written in total.
5. Extract the top 50 users based on total score.
6. Extract the top 50 users based on average score.
7. Extract the top 50 users based on accepted/answered ratio.
8. Extract the top 50 users based on Z-scores that normalize the ask/answer ratio.
9. Extract the top 50 users based on the total answers the persons they answer have posted.

Note that analysis 8 could be extended to cover all the other attributes as well, and for different levels of recursion. To approximate the second case we use the page rank algorithm [4].

To evaluate how similar two sequences are, we use two different measurements. First, we determine how many elements the two have in common. This is calculated using set intersection. Second, we compute the similarity by measuring how often the two sequences agree by calculating the number of agreements multiplied by 2 and divided by the total length of both sequences. Table 1 shows the results of these computations.

We find that analysis 1-5 and 9 often result in similar results. Depending on which pairs we compare, the results vary, but these often include approximately 75% or more of the same experts, in varying positions. Analyses 1 and 2 agree on 47 out of 50 users and are similar to approximately 38%. The latter may sound low, but in many cases, the two sequences only disagree on specific order. For example, what one analysis considers to be the fourth expert might be considered the fifth expert by the other, and hence, there is no match. So, 38% can be considered a high similarity, combined with the 94% match of total users. Analyses 3 and 4 result in the most similar sequences (44%), with 90% match between total users.

Analyses 6, 7, and 8 differ not only from 1-5 and 9, but also among themselves. These are not similar (0%) with no users in common. For analyses 6 and 7, this is due to users that have answered few questions but still have accepted answers or high scores. Analysis 8 is also affected by the actual number of posts; a user

that has posted some answers and no questions will score higher than a user that has posted many answers and some questions (since no questions is further from the equal balance assumed by the Z-score).

There is little to no change if we remove extreme values from analyses 6 and 7, such as users with 1 post with a very high score or users that have answered 1 question and that answer was accepted. The top 50 in analysis 1 have an average score of approximately 5, while the top 50 average scores are all above 100. We need to remove users with less than 1,000 posts to get average scores below 10 in top 50. Similarly, there exist a number of users with a single accepted but few answers in total, so the ratio in the top 50 is again quite high compared to the ratio of the frequent answerers. However, there are also a number of extreme cases, as discussed in Section 2.1.

Analysis 10 applied an unmodified page rank algorithm to the network with users that answered users. The result of this algorithm should be somewhat similar to analysis 9, which is a trivial version of the algorithm. The result of the page rank is similar to analyses 1-5 and 9.

3.3 Topics and Experts

If we assume that we can identify experts using one or more of the approaches defined in the previous section, we can investigate whether there exist different experts for different topics. As discussed in Section 2.2, each question is tagged with 1 to 5 distinct topics. Since there are more than 30,000 topics, and many of these are highly related, we attempt to identify experts in selected topics and groupings of topics. We use the following groupings and selections:

1. The top 5 topics with respect to number of posts per month
2. The top 5 topics with respect to how many topics they have co-occurred with
3. 4 groups of topics found by a community finding algorithm that approximately covers:
 - a) Microsoft Technologies (e.g., asp.net, ado.net, c#)
 - b) Mobile Development (e.g., iOS, Android)
 - c) Web Development (e.g., Flash, Javascript, CSS)
 - d) Systems Development (e.g, JVM, compiler, junit)

The top 5 topics per month, based on the full data set are: Android, Java, Javascript, c#, and PHP (in no particular order), and the top 5 topics with respect to co-occurrence are: c#, Java, PHP, Javascript, and jQuery. Given the large overlap, we include 4 additional topics based on rank: iPhone, .net, c++, and asp.net.

We find that there are some experts that have expertise in several fields. If we compare the 4 groups of topics, we find that Microsoft Technologies and Mobile Development are similar to 10% and 14 users are among the top 50 in both. There is also some overlap between Microsoft Technologies and Web Development, and Mobile Development and Web Development, that both share 1 expert among

the top 5, which hold the same rank. The topic groups overlap with some of the popular topics as well. For example, it is no surprise that there is some overlap between Microsoft Technologies and `c#`, `.net`, and `asp.net`. Similarly, we find overlap in expertise between `c++`, Systems Development, and Java. If we consider iPhone, a topic that has no obvious connections to any of the other topics, we find overlaps with Javascript. However, we find very little overlap between Web Development and web technologies such as Javascript, PHP or jQuery, for example. This could suggest that our method is flawed.

If we compare the experts within each topic and topic group to the global list of experts determined by page rank in the previous sections, we find that Microsoft Technologies, `c#`, `.net`, and `c++` overlaps to some degree. The largest overlap is with `c#` where 5 experts are among the top 50 of both, and the sequences are 4% similar (which suggests that there is at least one position that matches). This is not surprising since `c#` and `.net` are among the most popular and common topics discussed on the site.

4 Discussion

We introduced 10 different analyses to find experts. Based on experiments on the Stack Overflow data set, we find that 7 of these analyses provide results that are similar to some degree. Given the large amount of data (millions of users and posts), it is not reasonable to expect perfect matches; however, the 7 analyses that are similar often share more than 50% of the users and about 20-30% of the exact positions. This is not surprising, since they all include the number of answers a user has posted or the number of different users a user has answered. Page rank is also based on the number of incoming links.

The remaining 3 analyses are interesting since each of these is completely different from the other 9. We expected these to be different, given that the ratio and average can be misleading for users with few posts or very high scores per post. Similarly, the normalized Z-scores will prefer users that have not asked any questions to users with many answers and a few questions. However, in the case of ratios and averages, it required a significant filtering (ignoring users that posted less than 1,000 answers) to find some similarity between these 2 and the 7 that are similar to each other. This strongly suggests that each of the 3 and either of the 7 provide different ways to rank experts, that is, we have 4 different ways to describe an expert. We can either look at the average amount of up votes, the ratio of accepted answers, how different a user's communication is from "normal" communication patterns, or the number of total posts.

If we consider experts within different topics, our results are far from conclusive. We can find some overlap in expertise between different topics, as well as some overlap between specific topic and the global expert ranking. If we consider 4 groups of topics that a community finding algorithm identified, we find that some cases display the expected overlap while others do not. This is most likely a result of our categorization of the topics. This is something that will require further investigation.

We do not have access to a “proper” measure of who is an expert on Stack Overflow. So, there is no way to evaluate which of the 4 types of analyses, if any, provides the best estimate of an expert. There is a reputation system, where helpful actions, such as answering questions or improving the site, increase a user’s reputation. If we rank the users by reputation, we find that this ranking is similar to 1-5, 9 and 10. This is expected, since reputation is based on activities on the forums, so users with a high reputation will most likely have answered more questions, for example.

5 Conclusions and Future Work

In this paper we investigate if communication patterns can determine expertise and if so, what patterns. We rely on a data set that contains 4 years of questions and answers from the Stack Overflow developer forum. We find 4 possible ways to characterize an expert; amount of activity, average score, ratio of accepted answers, or how different the user’s communication is compared to “normal” communication patterns. We find that the four produce different results, but since we have no correct answer, we cannot determine which of these is the best or most correct. We can relate our findings to the reputation system on Stack Overflow, but this system favors activity and thus any characterization that relies on amount of activity will appear most correct.

We also compare experts within different topics to the global list and find that there is some overlap between similar technologies, such as `c#` and `.net`, for example. We do find inconclusive and contradictory results when we apply a community finding algorithm and attempt to find experts within groups of topics. Experts on Microsoft technologies overlap with for example `c#`, but experts on Web Development does not overlap with Javascript and jQuery. This is most likely due to our naive and simple approach to topic grouping, but the issue requires further investigation. We would also like to investigate if there are any topics with a reasonable amount of questions and answers that have very similar or identical groups of experts.

It would be interesting to perform a study of actual experts on Stack Overflow and compare these to our findings. This is made somewhat difficult given that the data set is to a large degree anonymized. However, the data is available on the site, and our algorithms could be adapted to work with the actual site rather than stored data.

References

- [1] A. Bacchelli. Mining challenge 2013: Stack overflow. In *The 10th Working Conference on Mining Software Repositories*, page to appear, 2013.
- [2] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

- [3] D. M. Boyd and N. B. Ellison. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1):210–230, 2007.
- [4] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1):107–117, 1998.
- [5] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph structure in the web. *Computer Networks*, 33(1–6):309–320, 2000.
- [6] A. Clauset, C. R. Shalizi, and M. E. Newman. Power-law distributions in empirical data. *SIAM review*, 51(4):661–703, 2009.
- [7] L. C. Freeman. Visualizing social networks. *Journal of Social Structure*, 1(1):4, 2000.
- [8] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, June 2002.
- [9] C. Haythornthwaite. Social network analysis: An approach and technique for the study of information exchange. *Library & Information Science Research*, 18(4):323–342, 1996.
- [10] P. Kollock. *The economies of online cooperation: gifts and public goods in cyberspace*. Routledge, London, 1999.
- [11] G. Robins. A tutorial on methods for the modeling and analysis of social network data. *Journal of Mathematical Psychology*, 2013.
- [12] G. Siganos, S. L. Tauro, and M. Faloutsos. Jellyfish: A conceptual model for the as internet topology. *Journal of Communications and Networks*, 8(3):339–350, 2006.
- [13] J. Zhang, M. S. Ackerman, and L. Adamic. Expertise networks in online communities: structure and algorithms. In *Proceedings of the 16th international conference on World Wide Web*, pages 221–230. ACM, May 2007.