

ネットワーク構造からみた Q&A コミュニティの分析

浅石卓真† 村山遼† 河村俊太郎† 芳鐘冬樹†† 鈴木崇史‡ 相澤彰子‡

† 東京大学大学院教育学研究科 †† 筑波大学大学院図書館情報メディア研究科 ‡ 国立情報学研究所

1 はじめに

Web が生活に浸透した現代社会において、人々がオンラインで情報を共有することの有効性は、今後ますます高まると思われる。本研究は、そのような情報共有の一形態である Q&A サイトに注目し、Q&A サイト内の質問者と回答者からなるコミュニティを分析する事で、オンラインでの情報共有活動に関する知見を得ることを目的とする。Q&A サイトの最大の特徴は、「誰でも何かを知っている」(everyone knows something) ことを、web を媒介して最大限に活用し、各自が得意な分野の質問に回答しあう環境を提供することである [1]。しかし、Q&A コミュニティ内でのそのような関係が成立しているかは、検討が必要である。実際には、多くの回答者から回答される質問者もいれば少数の回答者からしか回答されない質問者もあり、多くの質問者に回答する回答者もいれば殆ど回答しない回答者もいると考えられる。このような、情報共有における質問者・回答者の偏りは、Q&A サイトの理念にも関わることだが、まだ十分明らかにされていない。

一方、Q&A サイト内でのコミュニケーションには、1) 知識交換：質問の解が一意に定まる質問とその回答、2) 相談：解が一意に定まらず、質問者に明確な回答ができない状況でのやりとり、3) 議論：解が一意に定まらず、質問者が自身なりの回答をもっている状況でのやりとり、の 3 タイプがあることが指摘されている [1] [4]。これらのタイプに応じて、質問者の回答の得られやすさや回答者の回答しやすさは異なるために、質問者・回答者からなる部分コミュニティのネットワーク構造にも何らかの差異があると予想される。そこで本研究では、これら 3 つのコミュニケーションタイプで、1) 質問者の被回答数の偏り、2) 回答者の回答数の偏り、を調査し、その結果を比較・分析する。

2 分析方法

2.1 分析対象

本研究では分析対象として、Yahoo!株式会社が国立情報学研究所を通して提供した、2004 年 4 月から 2005 年 10 月にかけての Yahoo!知恵袋データを用いる。Yahoo!知恵袋内の中分類と小分類の全 286 カテゴリを、実際の質問文と回答文をもとに発表者の主観で前述の 3 タイプに分類し、質問者と回答者が 1000 人以上いる 128 カテゴリを実際の分析対象とした。各タイプのカテゴリの例は以下の通りである。

- 知識交換 (全 61 カテゴリ): パソコン・周辺機器、数学・サイエンス、祭りと年中行事
- 相談 (全 49 カテゴリ): 就職活動、結婚、子育ての悩み
- 議論 (全 18 カテゴリ): 政治・社会問題、アダルト、サッカー

2.2 分析手順

本研究の分析手順は以下の通りである。最初に各カテゴリで、質問者と回答者をノード、回答者から質問者への質問-回答関係をエッジとする有向グラフを作成する。次に、各カテゴリのグラフの入次数中心化傾向 (indegree centralization) と出次数中心化傾向 (outdegree centralization) [2] を算出する。本研究ではこれを質問者の被回答数の偏りと回答者の回答数の偏りの指標とする。グラフの中心化傾向は以下のように求められる。 n 人から構成されるネットワークで、ノードの中心性を $C(p_i)$ ($i = 1, 2, 3, \dots, n$)、中心性の最大値を $C(p^*) = \max C(p_i)$ とすると、各ノードの中心性の最大値からの偏差の総和は $\sum_{i=1}^n [C(p^*) - C(p_i)]$ となる。グラフの数学的な性格上考えられる偏差の最大の総和を $\max \sum_{i=1}^n [C(p^*) - C(p_i)]$ とすると、グラフの中心化傾向 C は

$$C = \frac{\sum_{i=1}^n [C(p^*) - C(p_i)]}{\max \sum_{i=1}^n [C(p^*) - C(p_i)]}$$

で与えられる [3]。なお、各ノードの中心性の指標は数多く提案されているが、本研究では、エッジ数 (被回答数・回答数) をそのまま中心性の指標とする度数中心性を採用した。最後に、各カテゴリのグラフの中心化傾向を基に、各タイプ間で中心化傾向の平均値を比較し、それぞれの特徴を分析する。

An analysis of Q&A communities in terms of network structure

† Takuma ASAISHI
† Ryo MURAYAMA
† Shuntaro KAWAMURA
†† Fuyuki YOSHIKANE
‡ Takafumi SUZUKI
‡ Akiko AIZAWA

Graduate School of Education, University of Tokyo (†)
Graduated School of Library, Information and Media Studies,
University of Tsukuba (††)
National Institute of Informatics (‡)

表 1: 各タイプのカテゴリの中心化傾向の平均値

	入次数中心化傾向	出次数中心化傾向
知識交換	0.028	0.056
相談	0.037	0.039
議論	0.056	0.056

3 分析結果と考察

図 1 に各タイプのカテゴリの入次数中心化傾向と出次数中心化傾向の散布図を、表 1 に各タイプのカテゴリの中心化傾向の平均値を示す。入次数中心化傾向でも出次数中心化傾向でも、各タイプとも一定の傾向があることが分かる。

入次数中心化傾向 (= 質問者の被回答数の偏り) は、議論 > 相談 > 知識交換の順に高い。議論タイプのカテゴリで被回答数の偏りが大きいのは、多くの回答者の関心をひく質問の質問者と、少数の回答者の関心しかひかない質問の質問者がいるためと考えられる。一方、知識交換タイプのカテゴリで被回答数の偏りが小さいのは、解が一意に定まるという質問の性質上、多くの回答者から回答される質問は現れにくく、質問者の被回答数も同程度になるためと考えられる。相談カテゴリでは、解は一意には定まらないがある程度は回答の方向性があるという、他の 2 つのタイプの間接的な性質の質問が多いため、被回答数の偏りも知識交換タイプと議論タイプの間接程度になると考えられる。

出次数中心化傾向 (= 回答者の回答数の偏り) は、知識交換タイプと議論タイプで高く、相談タイプで低い。知識交換タイプのカテゴリで回答数の偏りが大きいのは、カテゴリの内容全般に知識があり多くの質問に回答できる回答者と、一部の質問に回答するだけの回答者がいるためと考えられる。同じく、議論タイプのカテゴリで回答数の偏りが大きいのは、カテゴリの内容全般に関心があり多くの質問に回答する回答者と、一部の内容の質問にのみ回答する回答者がいるためと考えられる。一方、相談タイプのカテゴリで回答者の回答数の偏りが小さいのは、投稿される相談内容の中で、回答者は各々関心のある内容の質問に同数程度回答しているためと考えられる。

4 おわりに

本研究では、Yahoo!知恵袋の各カテゴリ内コミュニティをネットワーク構造として捉え、質問者の被回答数の偏りと回答者の回答数の偏りを調査し、コミュニケーションタイプ別のカテゴリ間でそれぞれの偏りの程度を比較した。その結果、1) 知識相談タイプのカ

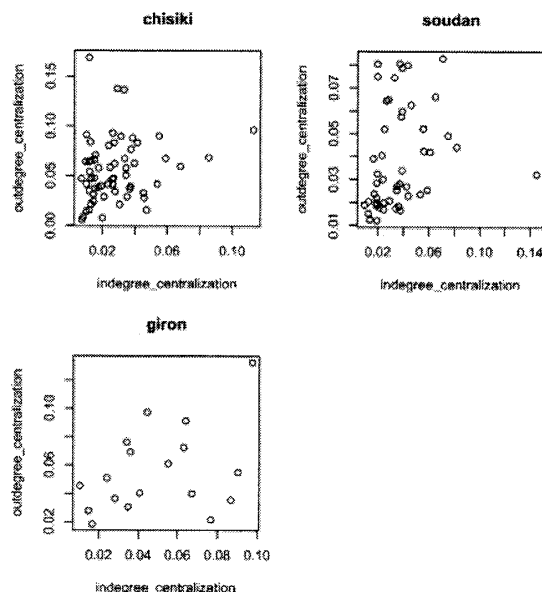


図 1: 各タイプのカテゴリの中心化傾向 (左上: 知識交換、右上: 相談、左下: 議論)

テゴリでは被回答数の偏りは小さいが回答数の偏りは大きい、2) 相談タイプのカテゴリでは被回答数の偏りは他の 2 タイプの中間程度で回答数の偏りは小さい、3) 議論タイプのカテゴリでは被回答数と回答数の両方で偏りが大きい、以上が明らかとなった。今後は、個別の質問者および回答者の中心性に注目することで、どのような質問者がより多くの回答者から回答を得ているか、また、どのような回答者がより多くの質問者に回答しているかについて分析を進めたいと考えている。

謝辞

本研究を実施するにあたり、Yahoo!株式会社が国立情報学研究所にて研究用に公開した Yahoo!知恵袋のデータを利用させていただきました。また、本研究は科学研究費補助金若手研究 (スタートアップ) の助成を受けました (研究課題番号: 21800087)。ここに謝意を表します。

参考文献

- [1] L.A. Adamic, J. Zhang, E. Bakshy, and M.S. Ackerman. Knowledge sharing and Yahoo Answers: Everyone knows something. In *Proceedings of the 17th international conference on World Wide Web*, pp. 665 - 674, New York, 2008. ACM.
- [2] L.C. Freeman. Centrality in social networks: Conceptual classification. *Social Networks*, Vol. 1, No. 3, pp. 215 - 239, 1979.
- [3] 金光淳. 社会ネットワーク分析の基礎, 第 1 章. 勁草書房, 東京, 2003.
- [4] 甲谷優, 川島晴美, 藤村考, 奥雅博. QA サイトにおける質問応答グラフの成長パターン分析. 日本データベース学会論文誌, Vol. 7, No. 3, pp. 13-18, 2008.