Conference contribution:
http://dx.doi.org/10.1145/3184558.3191631
An Evaluation of Performance and Competition in Customer Services on Twitter: A UK Telecoms Case Study

Nabeel Albishry  
University of Bristol  
United Kingdom  
n.albishry@bristol.ac.uk

Tom Crick  
Swansea University  
United Kingdom  
thomas.crick@swansea.ac.uk

Theo Tryfonas  
University of Bristol  
United Kingdom  
theo.tryfonas@bristol.ac.uk

Tesleem Fagade  
University of Bristol  
United Kingdom  
tesleem.fagade@bristol.ac.uk

ABSTRACT

With an increasing number of consumers using social media platforms to share both their satisfaction and displeasure about the products and services they use every day, organisations with a customer service focus are recognising the importance of rapid – and genuine – online engagement with their customers. In turn, consumers increasingly judge organisations on the quality of customer service and degree of responsiveness to online queries. This paper presents an extensible framework for evaluating direct engagements of customer service teams with customers on Twitter. Furthermore, this framework provides the capability to measure and analyse indirect engagement with industry sector rivals, especially their patterns, frequency and intensity. By applying graph analysis to these Twitter interactions, our framework generates various analytical measures and visual representations, exemplified through a case study based on seven major UK telecoms companies. With a dataset consisting of 15,000 tweets and 3,500 user profiles, the results provide sustained evidence for indirect engagements between business rivals, with customer queries acting as a trigger for intense competition between companies based in the same industry sub-domain.

CCS CONCEPTS

- Human-centered computing → Social networks; Social network analysis; Social networking sites: Information visualization;  
- Applied computing → Business intelligence;  
- Computing methodologies → Information extraction;

KEYWORDS

Customer services; reply chains; graph construction; social network analysis; Twitter; social media

1 INTRODUCTION

The online news and social networking service Twitter has become one of the most popular social platforms for a variety of demographics across the world. It provides a rich, constantly updating, corpus of big social data to study a range of complex socio-cultural issues, from life event detection [5] and identifying multilingual communities [2], through to sentiment classification [4] and providing deeper insight into personality and behaviour [17]. Unsurprisingly, Twitter is increasingly being used by organisations to communicate with their customers, due to the fast and convenient medium of engagement [15], using a variety of sophisticated human and automated approaches [25, 27]. In 2016, a survey was carried out on 5,450 people who follow small or medium-sized enterprises (SME) on Twitter [24]; the key results show that 83% of people that received a reply felt better about the SME, and 68.7% have made at least one purchase from an SME because of Twitter.

The medium can thus serve as an indicator to underlying issues of performance, management and even strategic matters [11]; in many instances, the majority of complaints deal with product and service-related issues [9]. Many studies have been conducted to explore aspects of customer services experiences in various business domains, such as travel and telecoms [14, 16, 23, 26, 28]. News agencies are not far from social media analysis, using it to uncover users’ interests so they can provide more focused contents [18]. While various domains have long applied network analysis techniques – such as for crime detection and prevention [19, 20] – only recently has work has been conducted to see how users relate to brands via network structures [8], how information shared by companies disseminate and their types [22], and what type of engagements from companies was found to be of effect on customers perception of the brand [13]. A common approach in conducting such studies has been to use sentiment analysis, mainly to measure consumer’s perception and satisfaction [1, 28].

However, the novel framework presented here aims to provide quantitative insights that can produce a more holistic view of customer service Twitter accounts and their interactions. Rather than focusing on individual posts and their sentiment, the framework helps in identifying complaint conversations that can be exploited...
by other business rivals, to be interrogated further by analysts or
decision makers. With the high volume of activity on Twitter, the
framework focuses on detecting possible key issues by using the
connected component feature of graph to identify problematic con-
versations for further analysis. Furthermore, by using streaming
and RESTful data, this approach can be applied to live data to catch
problematic conversations before they reach certain thresholds.

The remainder of this paper is organised as follows: in Sections 2
we introduce the case study of the UK telecoms sector; in Section 3
our methodology for this project; Section 4 presents the results
and key visual representations; Section 5 provides the main discus-
sion; Section 6 concludes the paper with a discussion of potential
extensions and wider application of this work.

2 CASE STUDY: UK TELECOMS SECTOR

Most UK households have access to both fixed broadband and a
smartphone, with consumers moving seamlessly between fixed and
mobile connections. This has been driven by the growing take-up
of superfast broadband services, with the proportion of UK house-
holds with fixed broadband increasing to 82% in 2017, with telecoms
services rising to 3.8% of total household spend. With the increasing
convergence of mobile and Wi-Fi connectivity, many customers in
the UK have switched from pay-as-you-go tariffs to pay-monthly
tariffs in 2016, and nearly two-thirds of mobile connections were
4G-enabled at the end of 2016. Consumers are also using these
networks more – average data use per fixed line residential broad-
band connection increased by 36% year on year to 132GB in June
2016, and average data use per mobile connection increased by
44% to 1.3GB [21]. The UK telecoms sector has continued to grow
rapidly over the past five years, with total sector revenues in 2016
of £35.6bn, including mobile retail revenues of £15.3bn [21].

There are four main fixed broadband network operators in the
UK: BT, Sky, TalkTalk and Virgin Media; alongside the incumbent
BT, alternative providers compete in the retail provision of fixed
services (including telephony and broadband), as well as other op-
erators using a variety of wholesale inputs purchased from BT and
resellers. There are currently four main mobile network operators
(MNOs) in the UK (with 2017 subscriber numbers): BT/EE (29.8m),
O2/Telefónica (25m), Vodafone (17.6m), and Three/H3G (12.01m); there
are also several mobile virtual network operators (MVNOs),
including Virgin Mobile (via EE, 3m), giffgaff (via O2, 420,000) and
Sky Mobile (via O2, 335,000) [10].

3 METHODOLOGY

The dataset contains tweets and related replies for seven well-
known UK telecoms companies: BT, EE1, giffgaff, O2, Sky, Virgin
Media and Vodafone. The choices were intended to represent compa-
nies of various sizes, history and range of services provided. While
a few companies only had one account on Twitter, some of them
have multiple accounts alongside the primary Twitter account; in
those instances, the dedicated customer services accounts were indi-
cated in the biography of the company’s other accounts. Therefore,
as the focus of the study is on customer services on Twitter, data
were collected from either the company’s primary account or its
dedicated customer services one (N.B. names throughout the paper
will refer to Twitter account handles rather than official company
trading names).

Inspired by the approach taken by Cogan et al. [7], this study
consists of two main steps: the data collection phases and the graph
construction. The data collection phase runs iteratively to obtain
reply chains, process them and store them in a database. Once the
data collection phase is completed, a large graph that includes all
reply nodes and edges is constructed to conduct the initial anal-
ysis. The NetworkX Python package [12] was used for the graph
construction, while Gephi [3] provided a range of tools for visuali-
sation.

3.1 Streaming

To ensure we were able to collect as much data as possible, the
data collection comprised of three steps. First, a stream endpoint is
opened to catch activities of accounts under investigation, those
accounts will be referred to as ‘CS’ (customer service) accounts.
The Twitter Streaming API2 is designed to return tweets created
by the user, their retweets, replies directed to their tweets, and
retweets of their tweets. However, the stream does not include
tweets mentioning the user, and replies/retweets by protected users.

3.2 Reply Chains

Returned statuses from the Streaming API may represent reply-to
statuses that have not been collected previously. It was found that
most missing statuses were either posted before the data collection
started, were mentions, or that the user account is protected. This
issue could have a significant impact on the quality of the analysis;
therefore, once statuses are returned from the stream endpoint,
their type is checked first (tweet, retweet, etc). If status is a reply,
the ID of the status to which it was replying is extracted from
in_reply_to_status_id. Then, using the extracted ID, we check if
the replied-to status has already been collected and present in the
dataset or not. If not, the REST API is then used to collect them.
This process runs recursively for newly-collected replies until no
further replies are available. Unavailable statuses are often results
from either deletion or protected accounts.

An analysis of changes on the graph after the second phase of
data collection shows that there were increases in the number of
nodes and edges by 43% and 62%, respectively. This increase in
connections has resulted in merging 176 conversations into others,
which improved connectivity of the graph and, thus, the accuracy
of the dependent analyses.

3.3 Graph Construction

The main data structure of conversations on Twitter suggests status-
to-status 3 relationship. In graph concept, statuses represent nodes
that are linked by directed edges. Therefore, the study follows a
graph construction approach in conducting analysis and produces
three graphs. First, base graph is constructed to capture structure
of the dataset, i.e. status-to-status. Then, from the base graph, it
generates two graphs: Users graph to examine the user-to-user

---

1BT’s acquisition of EE was completed in 2016; the merger did not materially increase
EE’s market share, but as a result of the merger EE’s usable spectrum share increased.
There have been no other significant new entrants or changes in market share since.

2https://developer.twitter.com/en/docs

3status’ refer to any type of post, tweet, retweet, reply, or quote
direct engagements, and Coexistence graph to uncover and examine indirect engagements amongst rivals.

3.3.1 Base Graph. Once the data are collected, a base graph is generated containing all replies and all related information. Nodes represent status IDs, while edges indicate replying direction; other information is added as attributes to nodes. The information used in this study are screen_name of the user, timestamp of the reply, text, and CS. The additional CS value is a binary digit set to distinguish accounts — it is set to 1 if the status belongs to one of the CS accounts, otherwise it is 0. This value is required to eliminate the need for user checks in forthcoming analyses. Figure 1 illustrates the conceptual base graph. As a reply can be directed to only one other status, no edge is expected to have weight value other than 1, and no reply status can have outdegree greater than 1. Nodes with 0 outdegree can be either a root node, or it is directed to unavailable statuses. On the other hand, indegree in this graph indicates the number of replies directed to the status node; hence, 0 indegree distinguishes leaf nodes. Special case nodes are those with indegree and outdegree equal to 0; these are isolated/ floating nodes and must be removed before we perform the analysis — these nodes do not benefit the analysis as they are not part of a conversation. Furthermore, they will be seen as connected component by themselves, which impacts upon the accuracy of results.

3.3.2 Users’ Graph. Because most of the analysis focus on relationships between reply posts, they were applied on the base graph. Nevertheless, to allow examination of the relationships between users, another graph is generated from the base graph. This process is carried out by iterating through edges linking reply posts, extracting users’ information, and constructing users graph accordingly. In the context of this study, only two attributes are used: screen names and ‘CS’ values. While nodes represent screen names, ‘CS’ values are attached to nodes as attribute. For edges, their weights indicate number of replies sent from origin node (sender) to target node (receiver); therefore, the user graph is directed. Applying this process on the example in Figure 1 results in the users graph in Figure 2.

To examine relationships between users, five network graph properties are measured. There were no special case nodes or edges in this graph, as observed in the base graph. For edges, their orientation indicate direction of replies, while weight reflects number of replies on the edge. Node indegree reflect number of users that have sent reply to the node, and outdegree indicates the number of users that have received reply from the node. Also, weighted measure of indegree and outdegree indicate total received and sent replies, respectively.

3.3.3 Connected Components. Reply conversations in the base graph are not interconnected, which implies that the graph actually consists of many subgraphs, or ‘connected components’. Since nodes include screen names, those components can be linked to CS accounts. Then, they are used to measure the size of conversation, their depths, and to identify shared conversations between the CS accounts. In the base graph, the number of connected components reflect the number of conversations. Therefore, in the base graph many components should be expected, depending on activity of the CS accounts and their audience.

To find conversations for a specific CS account, the search run through all components; in each component, the process iterates through nodes and examine the name attribute. Once a match is found, the search process stops and the identified component is either analysed on the fly, or returned for further analysis. Additionally, some of those components will be used to construct the coexistence graph, as covered next.

3.3.4 Coexistence Graph. As mentioned previously, this graph aims to measure the indirect engagements amongst the CS accounts. Therefore, all components in the base graph are processed in turn to find which CS accounts appeared. For each component, name attributes for nodes with CS value equal to 1 are extracted. If more than one CS account was found, those accounts are used to create nodes and edges for the coexistence graph. Thus, nodes are generated from CS names, and edges indicate common conversation between linked nodes. Subsequently, if the edge already exists, its weight is increased. For an illustration of this process, see Figure 3 and Figure 4 to exemplify three common components and the resultant coexistence graph.
4 RESULTS

4.1 Delay
Calculating delays is important to provide insight on the performance of the various CS teams. As reply nodes in the base graph include timestamp attribute, measuring delay is achieved by calculating time differences between end nodes on each edge. Table 1 shows key statistics for CS account delays; interestingly, @skyhelpteam was found to have an average delay of 45.04 hours, although the rest of the CS accounts’ delay ranged between 1.14 and 3.34 hours.

4.2 Interaction and Users
To measure interaction amongst users, user-user graphs were built from the base graph.; the resultant graph contains 3,521 user nodes and 5,938 edges. Although edges in the base graph cannot have a weight greater than one, edge weight in the users graph includes all replies from one user to another. Therefore, the number of nodes and edges in the users graph is lower that those in the base graph.

Properties of the users graph are presented below in Table 2; the table show that @virginmedia received that highest number of replies from 866 users with an average of 3.05 per user. Also, the same account scored highest in the number of recipients. The difference between indegrees and outdegrees shows that apart from @o2, all accounts have outdegrees bigger that their indegrees. Additionally, the total number of sent replies is found to be more than the number of received replies; this may reflect that those replies were directed to non-reply posts.

4.3 Conversation Components
As discussed in the methodology, each connected component in the base graph represent a conversation component that includes related replies. In this dataset, there were 3,289 conversation components with various number of replies. Observations of their sizes shows that the smallest component consists of one post, while the largest component contains 81 posts. The number of one-post components was 102, and they were all found to belong to CS accounts. Examining those singular components revealed that they were either original tweets that have not received replies, or replies to unavailable statuses. As covered earlier, unavailable replies are those that could not be captured due to a deletion or the posting account being protected. Because they do not have any length, and hence do not represent conversation, single-node components have been excluded from forthcoming analyses.

Additionally, the majority components were found to have two nodes. Those components were 1,188 and the orientation of their edge’s direction suggest that most of these communications were from CS accounts and directed to customer’s post. However, 25 of those conversations were initiated by customers. As they are two-node components, it seems that those posts have not been
answered by the relevant CS account. Although other means of communications could have been used, such as direct messages, there were no visible sign of further interaction.

### 4.4 Component Size and Longest Path

It is important to note that the size of connected components does not necessarily reflect length of conversations, although there is a strong correlation between size of component and length of its longest path (0.88). As can be seen in Figure 5, many components measures are positioned in a near-perfect diagonal line; interestingly, the longest path in the biggest component (81 nodes/posts) was only 1.

![Figure 5: Size of components and their longest paths](image)

To illustrate properties of connected components, the largest 20 components were chosen for visualisation, as shown in Figure 6. The findings show that components with very high variations in indegree amongst their nodes mostly originate from CS accounts. An example of this claim is illustrated by the three big components in the figure; when observed, they were found to featuring advertising tweets that had received too many replies from Twitter users. For example, the root node in the biggest component was a post by @o2 that has indegree of 80, and all connected nodes have indegree of zero, i.e. they were not answered. On the other hand, the longest path component was ranked the third biggest component. It was found with a single leaf, and all other nodes along the path were found with indegree=1 and outdegree=1, forming what we call a simple chain, uniquely coloured in Figure 6. Additionally, 15 of those components were found to have originated from customer accounts, and they all take a semi-simple chain as they feature some branches.

Generally, simple chains can be identified where the number of edges equals length of the longest path in component. Simple chains account for 80% of the connected components in graph, of which 47% were found with the length of 1. This is in agreement with the results of connected component sizes presented earlier. Finally, Table 3 presents statistics on chains of individual CS accounts.

<table>
<thead>
<tr>
<th>Name</th>
<th>count</th>
<th>max</th>
<th>min</th>
<th>mean</th>
<th>stdev</th>
</tr>
</thead>
<tbody>
<tr>
<td>btcare</td>
<td>388</td>
<td>19</td>
<td>1</td>
<td>3.46</td>
<td>3.22</td>
</tr>
<tr>
<td>ee</td>
<td>324</td>
<td>9</td>
<td>1</td>
<td>1.98</td>
<td>1.43</td>
</tr>
<tr>
<td>giffgaff</td>
<td>147</td>
<td>12</td>
<td>1</td>
<td>2.39</td>
<td>1.98</td>
</tr>
<tr>
<td>o2</td>
<td>216</td>
<td>11</td>
<td>1</td>
<td>2.53</td>
<td>2.26</td>
</tr>
<tr>
<td>skyhelpteam</td>
<td>252</td>
<td>15</td>
<td>1</td>
<td>2.22</td>
<td>1.99</td>
</tr>
<tr>
<td>virginmedia</td>
<td>959</td>
<td>68</td>
<td>1</td>
<td>3.78</td>
<td>4.46</td>
</tr>
<tr>
<td>vodafoneukhelp</td>
<td>248</td>
<td>39</td>
<td>1</td>
<td>2.58</td>
<td>3.29</td>
</tr>
</tbody>
</table>

Table 3: Summary statistics on chain length for CS accounts

### 4.5 Common Components and Coexistence

As covered earlier, connected components in the base graph represent individual conversations. Therefore, those components were utilised to uncover indirect engagement amongst CS accounts. As reply nodes in the base graph include screen name of user, it is possible to identify those components featuring more than one CS account. For each connected component, names in reply nodes are checked if they belong to CS account or public. Components with more than one distinct CS name are then marked as common component. The results show that there were 39 common components in total; 38 include two CS accounts, and one includes three accounts. The graph presented in Figure 7 shows those components, with each CS account given a colour code for identification as the legend clarifies.

To explore these relationships further, coexistence graph was constructed, as clarified in Section 3.3. Edges in this graph are undirected and their weights indicate frequency of CS accounts appearing in same conversation. The resulted graph is shown in Figure 8, where node size is proportional to its degree to indicate how many other CS accounts the node has coexisted with, while darkness of node reflects weighted degree to show the total frequency of coexistence for the node.

The first observation on the graph is that @giffgaff account was not found in any common conversation. In contrast, @o2 was the only account that have shared conversations with all other CS accounts, while @vodafoneukhelp was found with the least common conversations. Nevertheless, weighted degree measure shows that @virginmedia was the highest in number of common conversations; 21 components, although its degree tells that those conversations were shared with only three other CS teams. The heaviest edge existed between @virginmedia and @btcare, followed by the edge between @virginmedia and @skyhelpsteam. Also, edges of @o2 show that it mostly appeared with @ee, and for @vodafoneukhelp it was @ee.

Additional observation on the coexistence graph provides insight into uncovering more specific service areas within the specific industry or sector. This was clear when the modularity of the graph was examined [6]; the result has unfolded into two communities, as shown in Figure 9. Also, industry knowledge regarding the following CS teams: @ee, @o2, and @vodafoneukhelp belong to a domain that is mostly focused on mobile services, while @skyhelpsteam, @virginmedia, and @btcare are mostly known to be focusing on landline and home internet services.
Furthermore, using a similar approach that was used in Section 4.1, the delay was measured in those components to evaluate if presence of competitor has influence on how quick CS team response. Interestingly, improvement in delays of 26%, 43% and 72% were observed for virginmedia, btcare and skyhelpteam, respectively. These improvements in delays, in addition to the frequency of common conversations for those companies, confirm the existence of online competition amongst them.

5 DISCUSSION

Initially, the performance of CS accounts and their popularity on Twitter were measured by an analysis of activity and users. From this perspective, @virginmedia was found to have the highest volume of posts, the least diverse in terms of type of posts (99.7% were replies) and with the highest number of customers served. The average delays of accounts ranged between 1.14 and 3.34 hours, apart from @skyhelpteam which was found with an average delay of 45.04 hours. This may indicate a management issue for the team, such as unclear social media strategy or staff resources.

Most CS teams have clearly specified working hours on their account page, apart from @giffgaff and @o2. Interestingly, these two accounts were found to have the lowest delay. Nevertheless, high availability, i.e. longer activity hours, was not found to significantly improve speed of reply to customers. For example, while @giffgaff was observed active for longer periods, @o2 was generally found to be faster to reply.

Although the data shows that no CS team has been in a direct engagement with a competitor, analysis of common/shared connected components has uncovered some form of competition amongst CS accounts. Particularly, in the case of @virginmedia and @btcare, the competition was clear and intense. In all cases, customers were found to be initiators of competing conversations by making use of the Twitter @-mention feature to bring different rivals into conversations. In contrast to phone, letter or email, complaints that are made on social media are open for the public to read and follow, and can be potentially reputationally damaging if not handled appropriately. Therefore, it was not surprising to see improvement in the speed of response in a few instances where competitors were included in the same conversation. This shows that with the openness of social media platforms, such as Twitter, customers may have more chance to obtain better deals or speedy resolution of their problems [9]. In turn, this approach of publicly-posting complaints add pressure onto CS teams to improve their social media engagement, especially when business rivals are included by customers [11].

Some of these issues are particularly pertinent for the mobile virtual network operators (MVNOs) – such as Virgin Mobile, giffgaff and Sky Mobie – who use the infrastructure of the main mobile network operators (MNOs), as presented in Section 2. While this
may not directly affect CS operations, especially as many customers are unlikely to be aware of whether their company owns any underlying network infrastructure, it may have operational and/or performance impacts on the virtual operators depending on their service-level agreements with the infrastructure operators.

6 CONCLUSIONS

The paper has introduced an extensible framework for evaluating customer service performance and competition between industry rivals on Twitter. We have presented methods on how network graph properties can be used to make increasingly sophisticated
evaluations, with the framework being tested on selected operators in the competitive UK telecoms sector.

Section 3 highlighted two important techniques that need to be applied prior to starting the analysis phase. First, the recursive reply chain data collection is vital to obtain accurate results. The importance of this stage stems from the fact that it fills the gaps and improve the connectivity of graphs. Second, construction of the initial graph from replies and the removal of floating isolated nodes. In constructing this graph, key information needs to be identified and attached as attributes to nodes. The information used in this study include post id, timestamp, screen_name, text and CS value. However, the framework could easily be extended to include other information, such as retweets.

A wider aim of this project is to show the importance of connected components in distinguishing users’ conversations, as well as analysing competitions and their key features. With the added value of modularity classes, competition analysis has helped in uncovering more specialist communities within the industry sector.

The presented framework could also be used by service providers to reflectively evaluate their social media accounts and interactions, as well as to generate insight into the activities of their key domain competitors; in this way, the presented methods in this study could be used to make real-time observations. Another application would be to identify gaps, competitions, challenges and opportunities in services that can be used in developing strategies for start-ups, for example. The approach could also be applied to other domains or contexts, such as non-profits, charities or the public sector. Also, it can be used for groups of users, such as celebrities and their direct and indirect engagements on Twitter. Moreover, with the emerging practice of signing a reply with a team member’s initials, the approach could also be applied to other domains or contexts, such as non-profits, charities or the public sector. Further, the approach could also be applied to other domains or contexts, such as non-profits, charities or the public sector.

ACKNOWLEDGEMENTS

This work has been supported by a doctoral research scholarship for Nabeel Albishry from King Abdulaziz University, Kingdom of Saudi Arabia.

REFERENCES


