

Telegram group quality measurement by user behavior analysis

Ali Hashemi , Mohammad Ali Zare Chahooki

Social Network Analysis and Mining

Presenter: Tuja Khaund



Introduction

- Telegram is a cloud-based instant messenger with more than 200 million monthly active users.
- Very popular and growing rapidly in countries such as Iran, Uzbekistan, Indonesia, Brazil, and Russia.
- **Features** : channels, bots, and supergroups.
- *Channels* are mostly used as news gateways, and the content is publicly available.
 - Only channel owners can post content on the channel.
- *Bots* offer a lot of possibilities and features to users.
 - Each user interacts with bots individually and gets personal responses accordingly
- *Supergroups* can have up to 100,000 members.
 - Every member of a supergroup can post content.
 - Administrators are able to report and ban other members.



Objective

- Telegram is the most popular instant messenger in Iran which attracted active marketers to Telegram environment
 - Telegram provides neither a marketing tool nor a search service.
- Information may be useful for many interested parties such as investors, marketers, and entrepreneurs.
 - 900,000 Persian channels and 300,000 Persian groups.
 - gives a brief insight into the Telegram environment to better understand further findings.
 - Extend their analysis to measure group qualities
 - Extract features from Telegram groups to represent them with numerical values.
 - Construct a dataset in which high-quality and low-quality groups are marked.



Data Collection

- Telegram Search - *IdeKav*
 - A search engine which covers more than 900,000 channels and 300,000 groups.
 - Search the content of the groups along with title and description.
 - Search is performed on an inverted index, and the ranking is based on TF-IDF scoring.
- Developed a cluster of crawlers to scrape data over a period of two months
 - Extracted 23 features from the dataset (*Example below*)

No.	Name	Description
1	AUC	Number of users who sent at least one message in the group in the respective 2-month period
2	AC	Number of administrators of the group
3	T	Title of the group (i.e., group name)
4	DT	The time the group has been discovered by us
5	AL	Average length of messages published in the group in the respective 2-month period (measured in characters)
6	CA	Average of number of messages forwarded from each unique channel to the group in the respective 2-month period
7	CC	Number of unique channels from which at least one message has been forwarded to the group in the respective 2-month period
8	CV	Variance of number of messages forwarded from each unique channel to the group in the respective 2-month period



Attributes of group quality

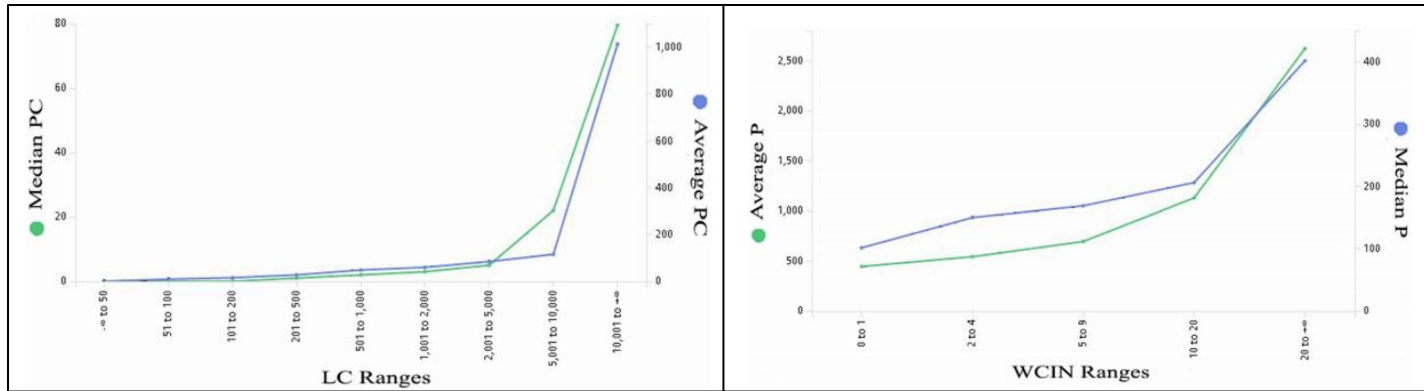
Attributes of a high-quality group	Attributes of a low-quality group
The group has a specific purpose. The topic is clear, and off-topic conversation is not welcomed	The group purpose is either unclear or just for passing spare time chatting. Participants chat about different topics
Spam and advertisement are rarely seen. Administrators kick spammers out	A lot of spam and advertisement are posted every day
Fake and multiple accounts are rare	Fake accounts are used to increase the member counter, so that new participants think the group is popular. Individuals join the group with multiple accounts, so that if one of their accounts got kicked out, they still have access to the group
Administrators constantly watch and moderate the group. Newcomers are first warned for unwanted behavior and then kicked out if they repeat it	Administrators do not care much about the content. It looks like they just want to have more members
The number of new daily posts is rational. An individual is able to read all new daily messages every day	Either thousands of new messages are posted every day, or the group is silent for weeks
Usage of bots is limited to tedious administration tasks like detecting and deleting spam content	Spam bots are used in the group. For example, the bot does not allow participants to post messages unless they add five new members to the group
It is unlikely to get private messages from other members of the group, unless you explicitly request for personal help in the group	There is a high chance of getting spam messages privately after joining the group. In other words, participants are fed to spam bots



Discussion

- Average message length (AL) is noticeably larger in high-quality groups compared to all others.
 - The average length of each message in high-quality groups of our dataset was 136 characters, while it was 48 characters for low-quality groups.
- Higher diversity of forwarded messages is seen in high-quality groups.
 - On average, messages have been forwarded to high-quality groups from 36 different channels, and to low-quality groups from 20 different channels.
- Low-quality groups have sent 8496 messages in total on average. The average for high-quality groups is 1445.
- Low-quality groups are the most active groups in terms of the number of total messages.
 - But, considering engagement ratio, high-quality groups go to the top of the list.
 - 1445 messages have been sent by 205 users in high-quality groups
 - 8496 messages have been sent by 570 users in low-quality groups.

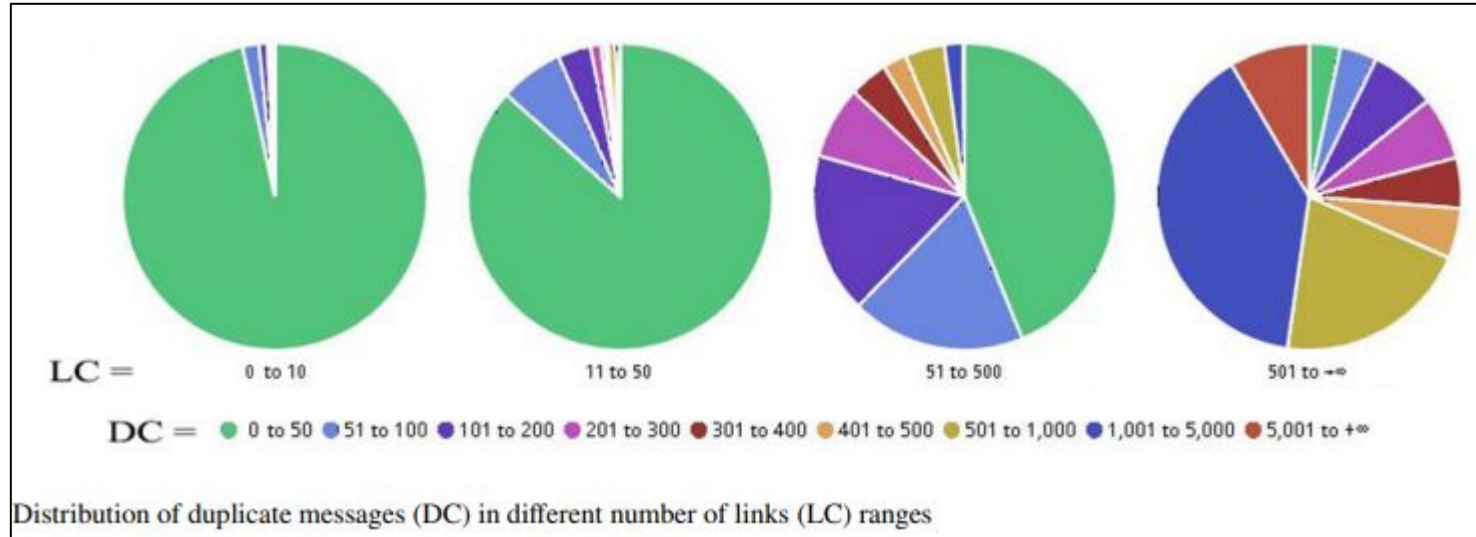
Further Analyses



Growth rate of the number of phone numbers (PC) by the number of links (LC)

Members' (P) growth by non-alphanumeric characters in the group name (WCIN)

Further Analyses





Conclusion

- High-quality groups tend to forward more messages from more different channels.
- In high-quality groups, more users are engaged in each discussion, but discussions are shorter.
- More phone numbers are shared in high-quality groups than others.
 - service providers share their phone numbers very often in high-quality groups.
- Users are more likely to join groups with more non-alphanumeric characters in the group name.
- The more a group tends to publish links, the more duplicate messages it will have.
- High-quality groups have longer messages, more diverse forwarded messages, less non-alphanumeric characters in the group name, and more user engagement.
- After censorship in Iran and Russia, users did not start using other messaging apps, but instead turned to VPN services to circumvent the block, rendering the censorship ineffective.



Thank you!

