Socio-spatial Characteristics of the Information Flow in Social Networks

Author: David Grob
Supervisor: Prof. Dr. Martin Raubal
Supervising Tutor: Florian Straub

Master Thesis, Autumn Semester 2012

Motivation

The concept of defining tie strength was introduced by Granovetter. The existing connections were classified as weak and strong ties. Strong ties link close friends and weak ones tie acquaintances. He propagated the strength of weak ties: In order to get a new job, the information from weak ties is most often successfully used. His explanation was that people who are connected with a strong tie have quite the same information space. Therefore, the diversity of answers is higher for weak ties and as a consequence it is more probable to get a job if this information is used.

Research Questions

- What about Granovetter's theory in an online social network?

- How can we characterise the information flow in an online social network in general?

- What role plays spatial context?

In order to be able to answer these research questions, the data of the social network has to be seen as multidimensional.
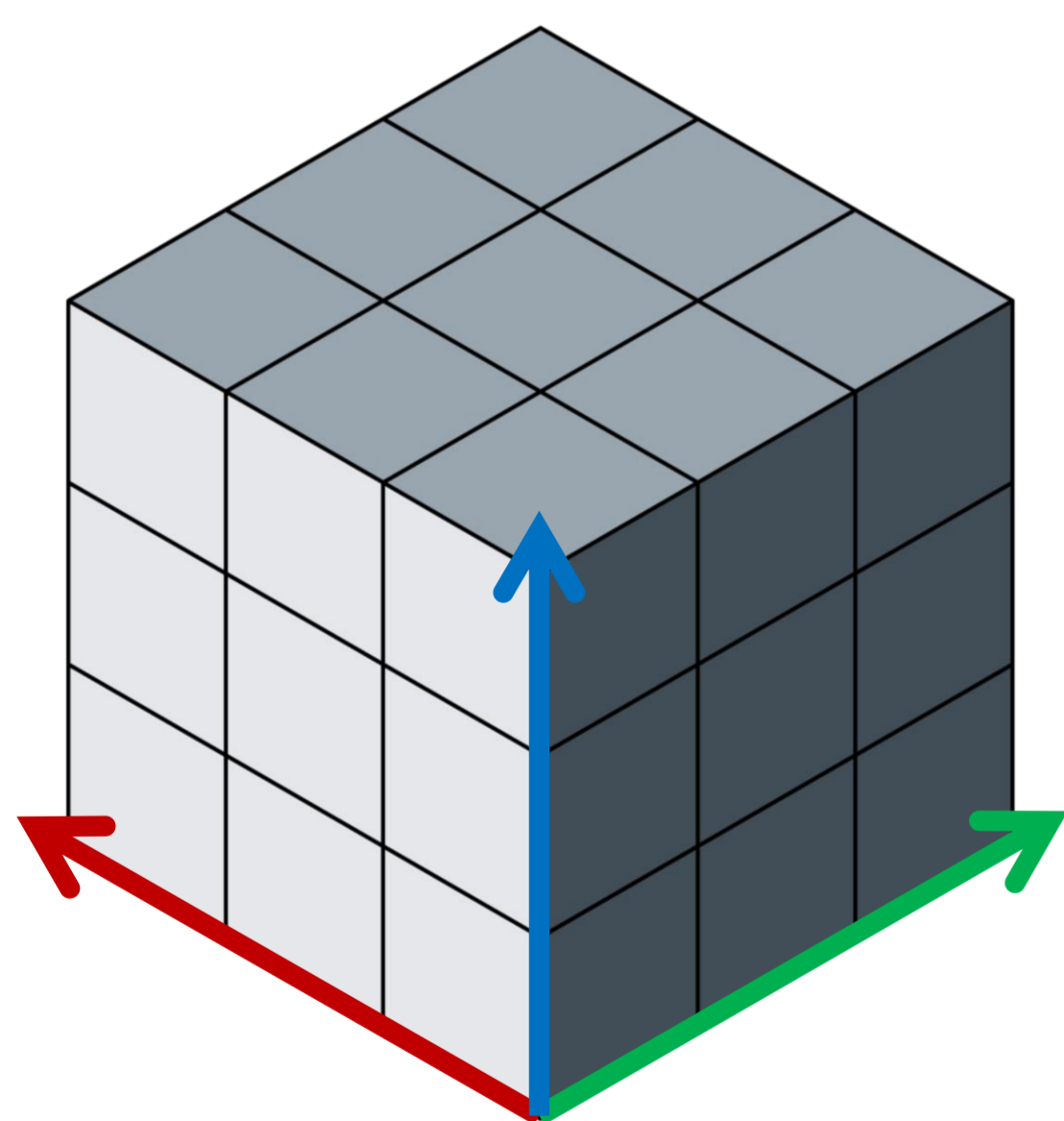

Figure 1: Visualisation of the dimensions (tie strength, distance and value of information)

The first dimension is the tie strength. Based on existing knowledge, the tie strength can be stated in a higher resolution than by Granovetter. The second one is the distance. The geographic distance between two persons has an influence on the existence of a tie and can affect the way of interaction. The third dimension is the value of information. Its magnitude describes how relevant and new the discussed topics between two users are.

Social Network Model

A social network can be modelled as a graph. The users of the social network represent the nodes and the edges describe the connections between them.
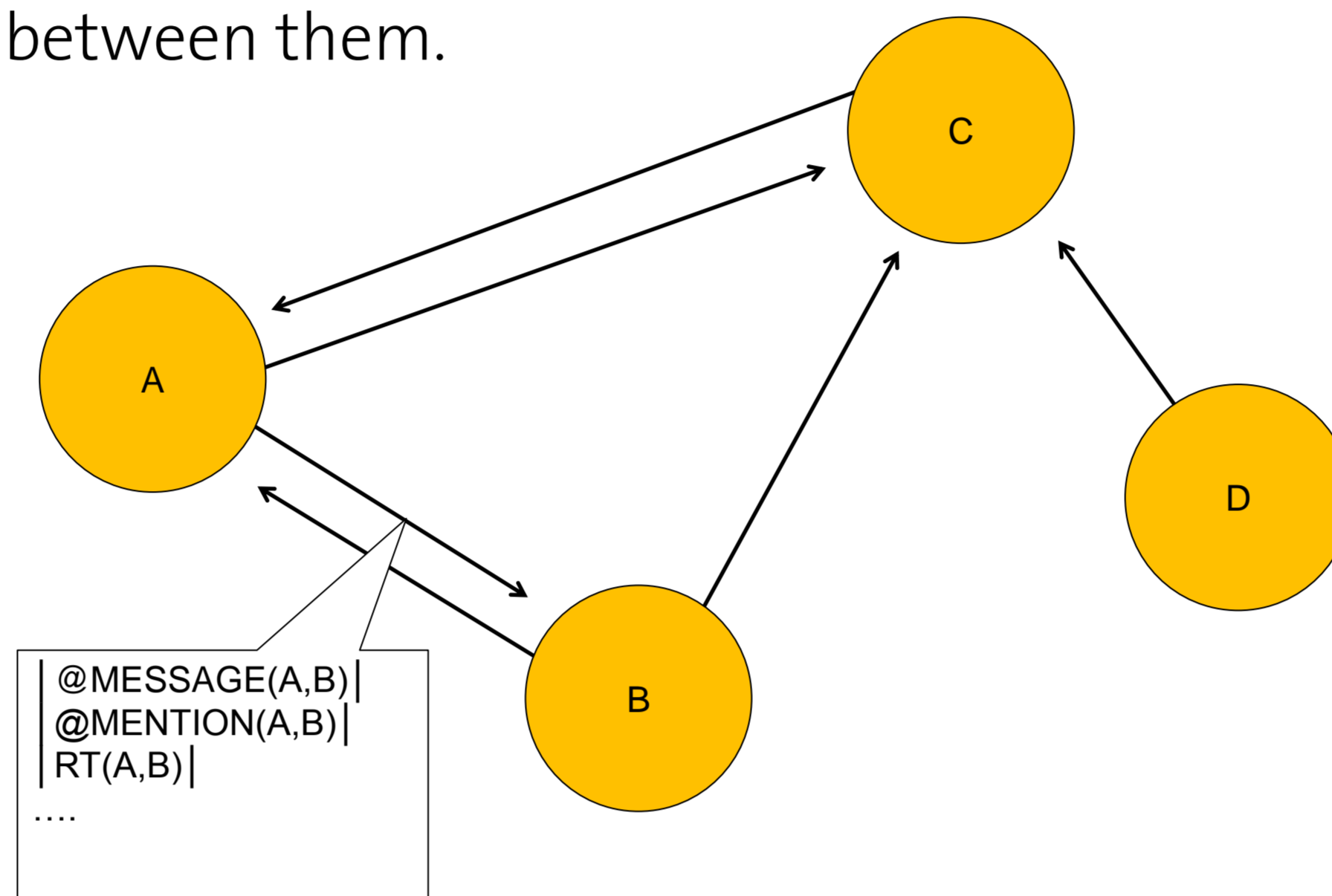

Figure 2: Social network model with attributes for the e.g. tie strength calculation

The used data is from the microblogging platform Twitter, which means that the length of the messages (tweets) is limited to a maximum length of 140 characters. All tweets can be read by the followers of the user who wrote them. A user can sent tweets to specific users (@messages). The information flow is used to calculate the tie strength and the value of information for every edge within the dataset. The distance dimension needs the users' home location which isn't always available. Users with no explicit coordinates are geocoded with the Google Geocoding API based on their indicated hometown. The geocoding result is combined with the user's geotagged tweets. The spatial communication density is visualised in the following figure:
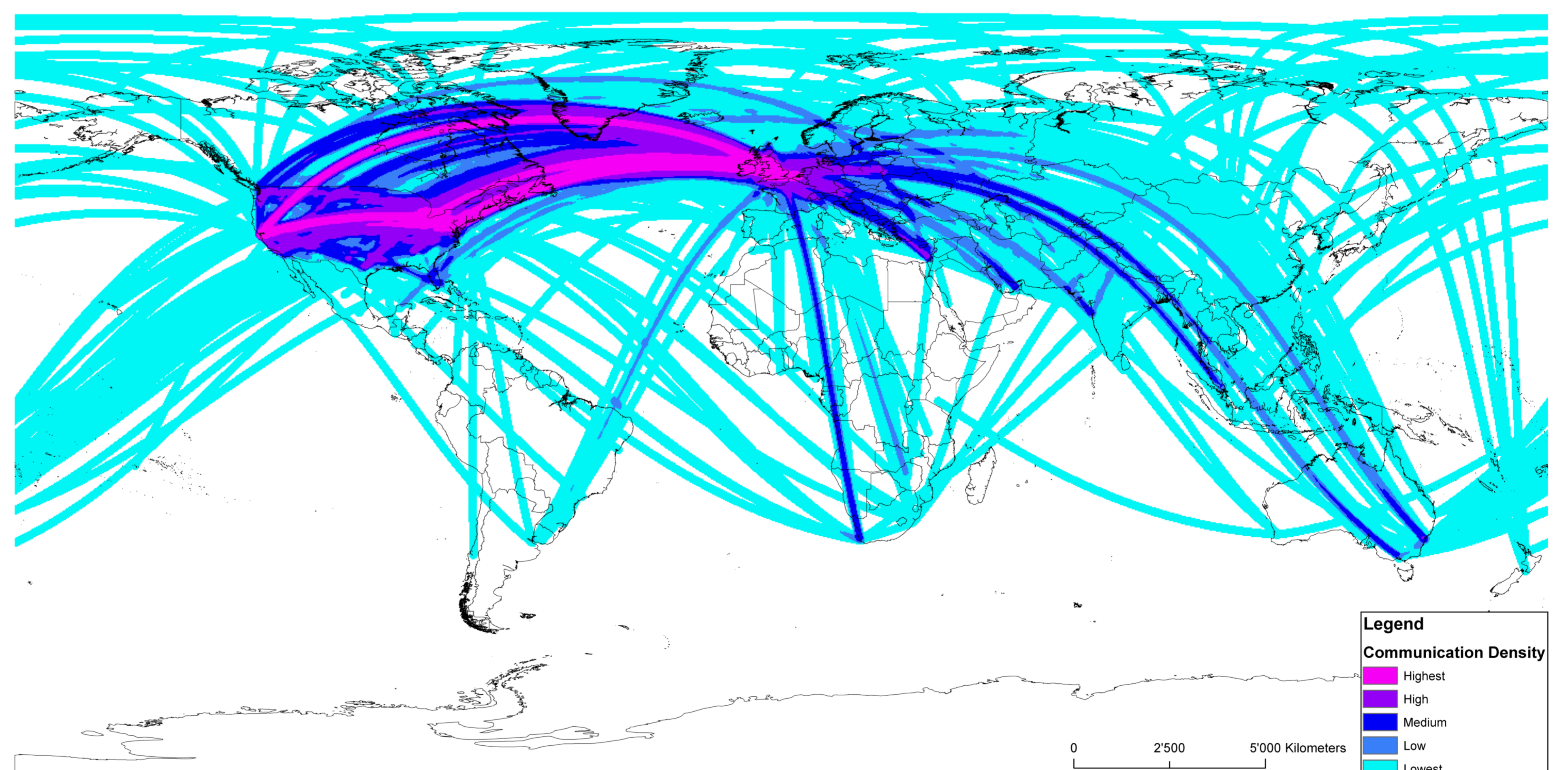

Figure 3: Communication density

Results

The average tie strength tends to decrease if the distance increases (see Figure 4). It reveals that on average better friends live spatially closer than not so good friends.
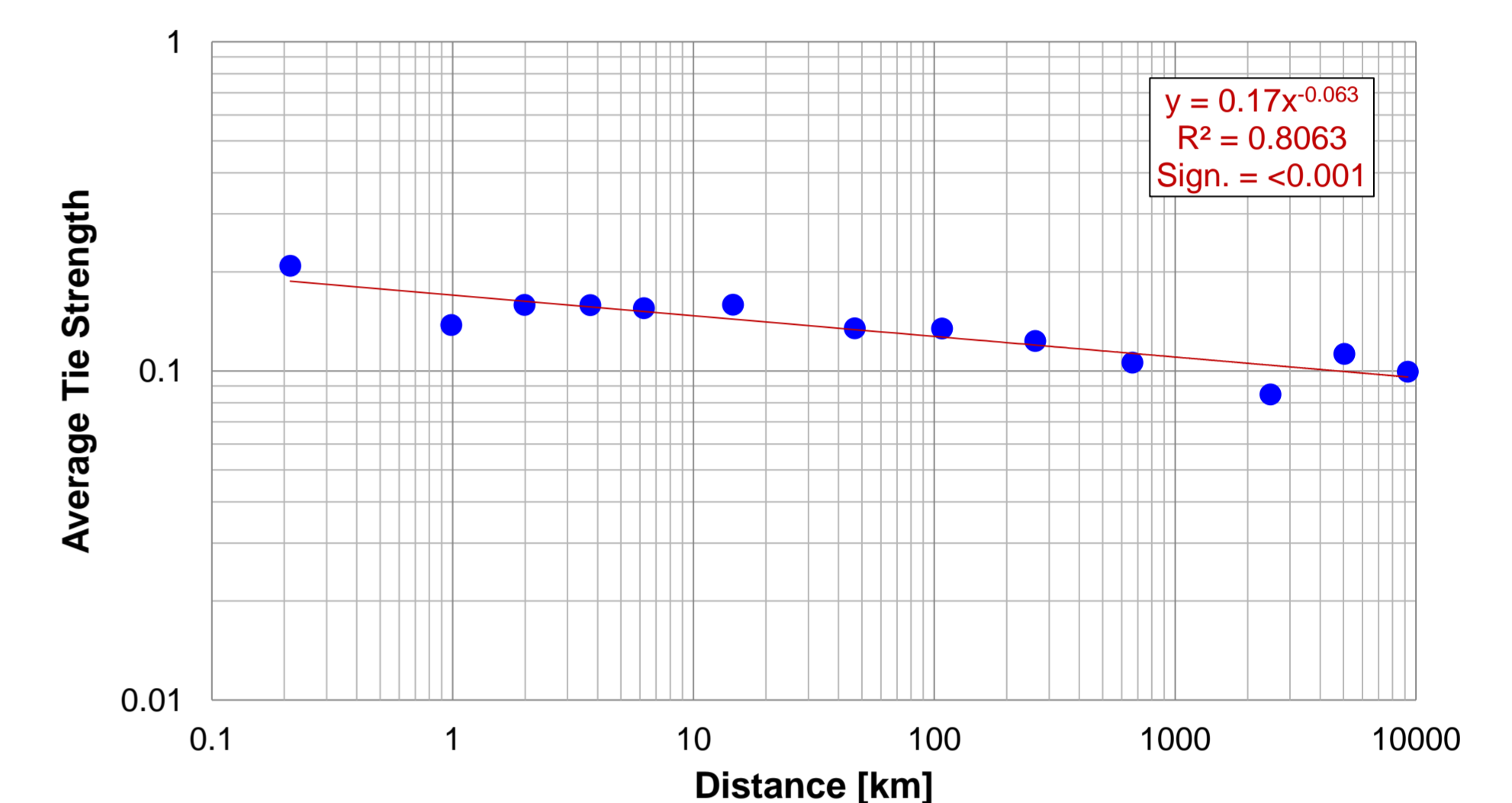

Figure 4: Distance versus tie strength

In Figure 5, the value of information is represented by the novelty. People tend to receive newer information over weak ties. Hence, the strength of weak ties is visible in the Twitter data. However, the value of information seems to be more or less independent of the distance.
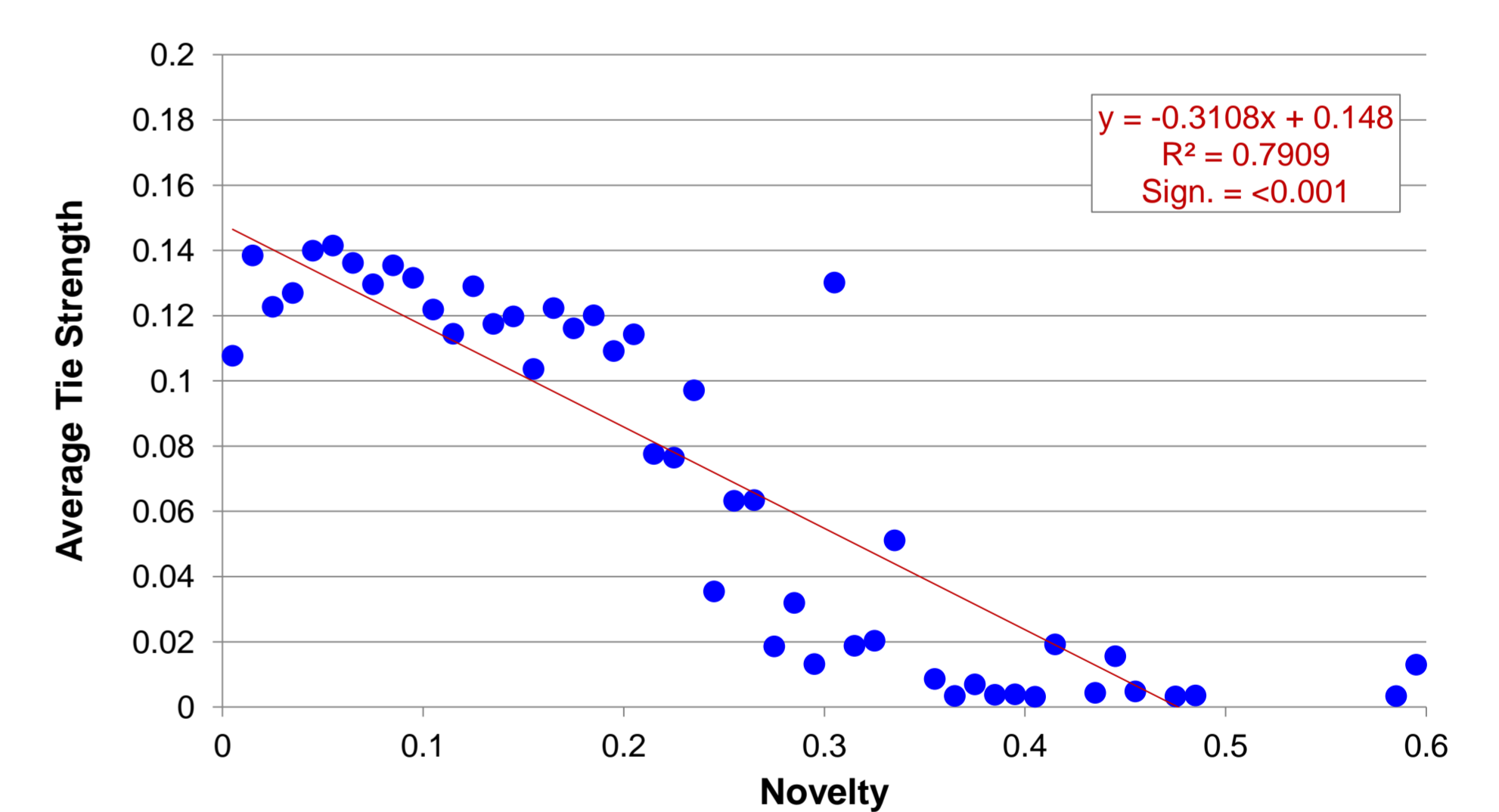

Figure 5: Novelty versus tie strength