# BDP Hackathon Report

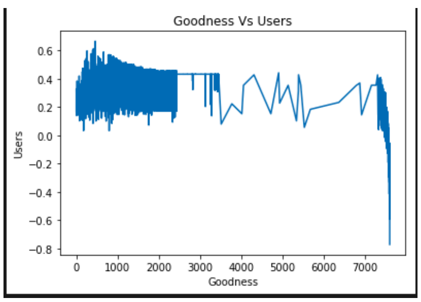
The code for the report and all data can be found at this link:

https://drive.google.com/open?id=1thVrXd-PZkHXeCuyzzE22ItW7YZGOYef

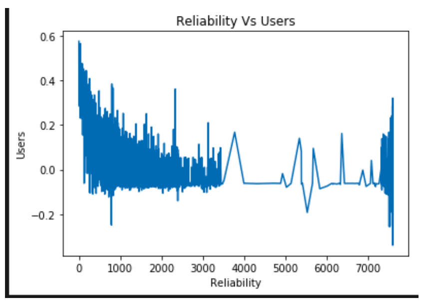
https://drive.google.com/open?id=1thVrXd-PZkHXeCuyzzE22ItW7YZGOYef

## Question - 1

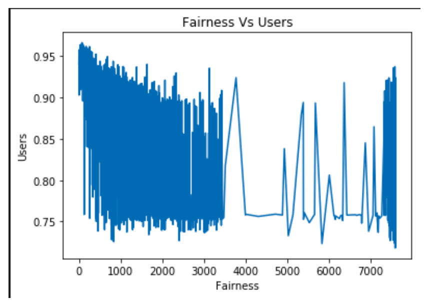
- The plots for fairness and goodness for all users are shown below.




*Goodness plot of all users.*




*Reliability plot of all users.*




*Fairness plot of all users.*

The fairness and goodness scores area unit computed supported the affiliation of nodes within the graphs weighted by the scores assigned  by users to every alternative mistreatment the algorithmic rule below, responsibleness is same as fairness as given within the paper.

**Input**: A WSN $G = (V, E, W)$
**Output**: Fairness and Goodness scores for all vertices in $V$
Let $f^0(u) = 1$ and $g^0(u) = 1$, $\forall u \in V$
$t = -1$
**do**
    $t = t + 1$
    $g^{t+1}(v) = \frac{1}{|in(v)|} \sum_{u \in in(v)} f^t(u) \times W(u, v), \forall v \in V$
    $f^{t+1}(u) = 1 - \frac{1}{2|out(u)|} \sum_{v \in out(u)} |W(u, v) - g^{t+1}(v)|, \forall u \in V$
**while** $\sum_{u \in V} |f^{t+1}(u) - f^t(u)| > \epsilon$ or $\sum_{u \in V} |g^{t+1}(u) - g^t(u)| > \epsilon$
**Return** $f^{t+1}(u)$ and $g^{t+1}(u)$, $\forall u \in V$

*Srijan Kumar et. al's algorithm.*

- Trust score is calculated by the following formula:

$$trustscore = fairness * goodness$$

  Using this formula, the top 10 users are:

```
791
828
932
861
790
963
829
978
831
833
```

  The bottom 10 users are:

```
7481
7533
7541
7538
7452
7457
7449
7468
7479
7456
```

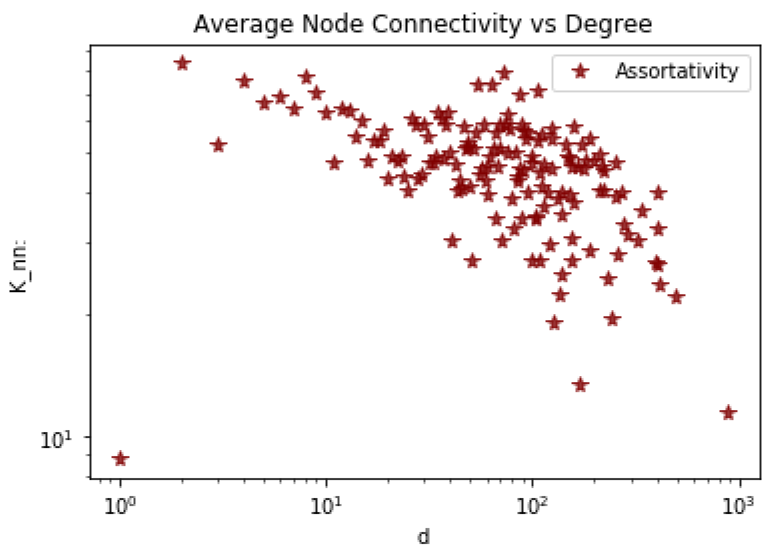
- The total length of strongly connected component is 3235 out of the total number of 3783 nodes in graph.

We can conclude that this is often the most node wherever the lifeline of the network exists with all the active users thereon. Rest users square measure isolated from the network. With this, we will calculate the overall active population on the network is **85.51%**.

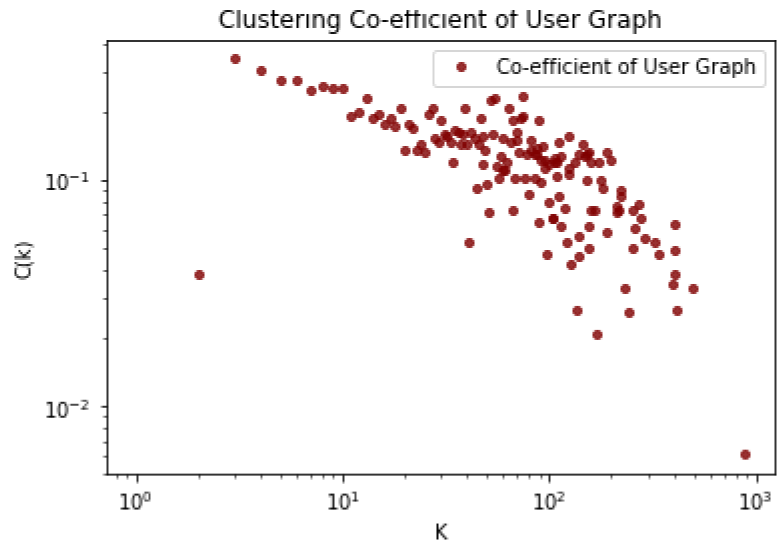
- According to Maine, person B can't be trusty as Person B has identical negative score as A, however with fewer users rating him which implies his average negative response with fewer ratings is that the same because the average rating response of A with higher users. several users' average positive and negative rating has given identical rating as few users' positive and negative rating.

- The other graphs include Clustering coefficient and assortivity of the network as shown below.

**Assortivity** on the opposite hand, is that the live of closeness of nodes of comparable degree in a very graph. we will see that, this live too is pretty high within the case of the network.

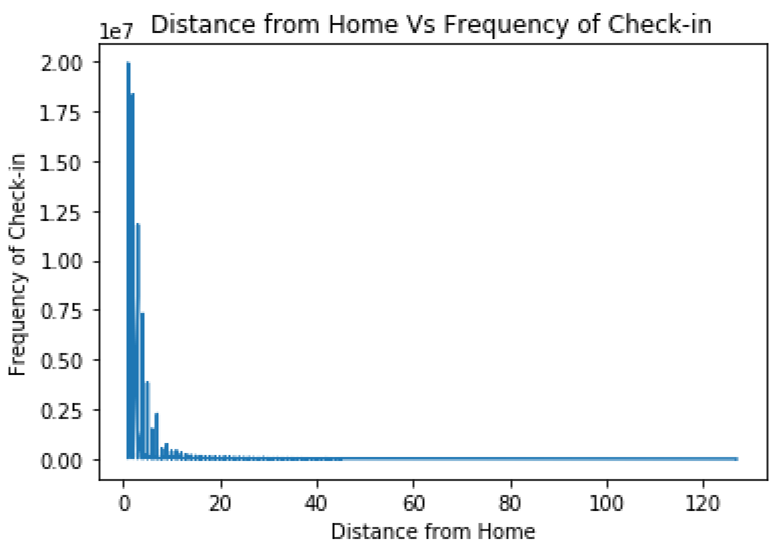



**Clustering constant** defines however shut square measure the users within the network graph. From the graph below we are able to conclude that almost all users have moderate or higher clump constant resulting in moderate to high property within the entire network.




## Question - 2

- The way the user tends to travel from home is given by:

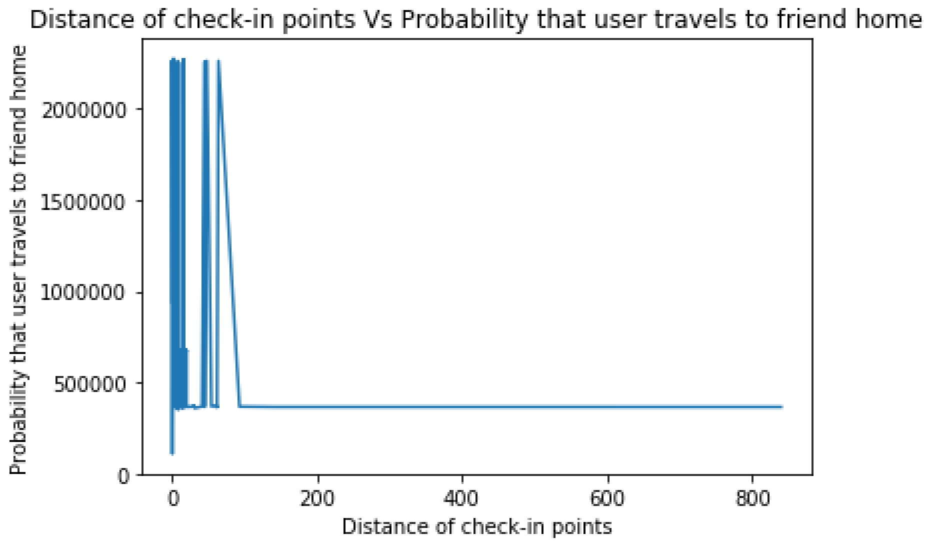



The plot represents an **Power Law.**
This shows inverse relation between the degree of nodes and therefore the range of nodes, in general.

The inferences which will be drawn from this is often that only a few users arrival at distant places from home. Moderate distance check-ins ar seen during a moderate range of users and really close to check-ins ar seen during a most of the users.

We will infer that folks typically arrival at places whose distances ar close, as a result of motion conveniences, and really few users like travel enthusiasts arrival at far-away places. This is also a sign of a user's movement around her/his house and reflects daily behaviour.

- The following graph shows the probability a particular user travels to a friend's home vs distance of check-in.




To  calculate this we tend to initial calculate the likelihood that a user visits an addict. we tend to divide the house segments of users in terms of circles of twenty five kms every. If a user checks-in this segments, then we tend to assume that the user is in his friend's territory. With this, we tend to calculate the count of friends overlapped at a distance d as:

$$probability = \frac{number\ of\ checkin\ in\ B25\ at\ distance\ \mathbf{d}}{total\ check{-}ins\ at\ distance\ \mathbf{d}}$$

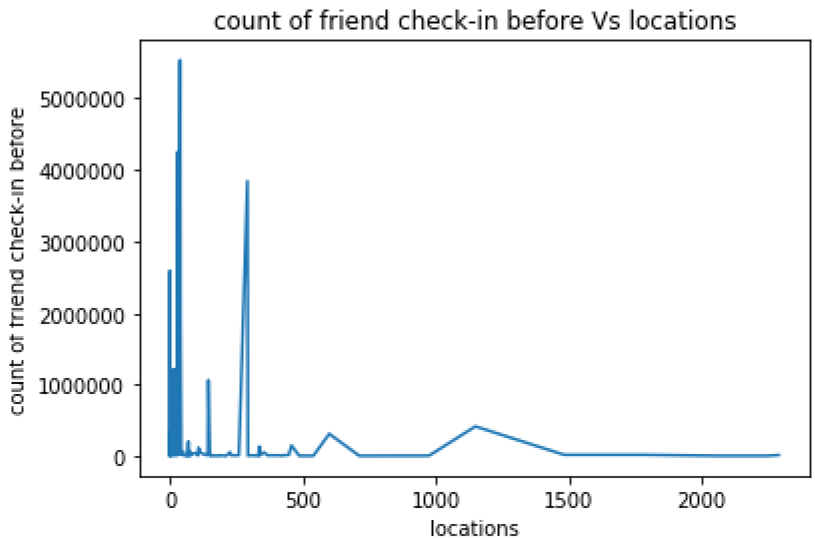
where $B25$ is the region of 25 kms of a friend's territory.

We have used the 25 km bounded range for two reasons:

- To overcome errors due to averaging home locations.
- Even check-ins in a nearby location to the friend's house is a meeting with the friend in most cases.

With this knowledge, we will infer that, either, the user's friends all live close or the close users have infact become the chums of the user. we will conjointly conclude the the user likes to travel to the present friends' location a lot of usually and doesn't prefer abundant to travel more distances because of motion conveniences.

- We had to see here the amount of check-ins  at a specific location of a devotee  of user before the user checks in. This makes United States perceive if the user visits a location visited by his/her friend before. we tend to see victimisation the graph below ,




that a great deal of users are influenced by their friends' locations and visited that locations. This once and for all proves that a users' friends will verify subsequent location of the users.
This is vital in investigation connected aspects.

- The limits of using friendship data in mobility is that:
    - Some users have no such connection between their friends' checkins and their check-ins. This disproves the fact that friendship can be used, since it depends on the  personality of the user i.e introvert, extrovert, etc.
    - This also depends on how active the user is on social media.
    - The general observation on social media is that acquaintances are also friends on social media. This combined with the fact that most check-ins are in nearby places with no actual checking/concrete proof that the friends are real/true friends of the user leads to false conclusions.
    - The same problem of whether social media does actually reflect the relationships in real life.