# Report, cv1

## DBLP network

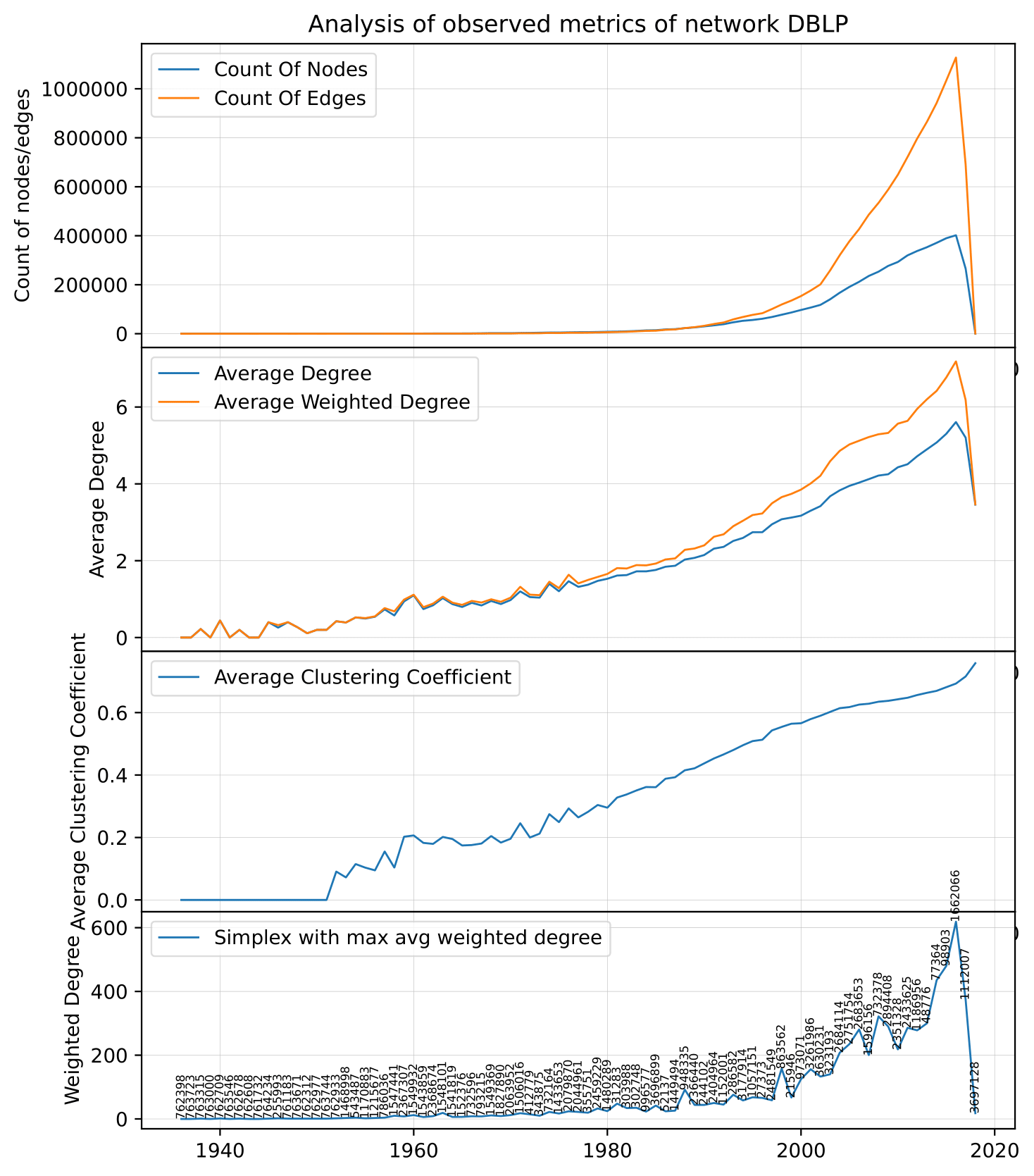
DBLP (Digital Bibliography & Library Project) is a well-known online repository that contains an extensive catalog of scientific works and publications in the field of computer science and related disciplines. One of the key features of DBLP is the ability to analyze collaboration networks within the field of computer science and the academic world in general. The collaboration network in DBLP connects authors and scientific works based on collaborations among individual authors.

The collaboration network in DBLP identifies connections (collaborations) between authors based on shared publications. If two authors have collaborated on the same scientific work (for example, as co-authors), they are linked within this network.

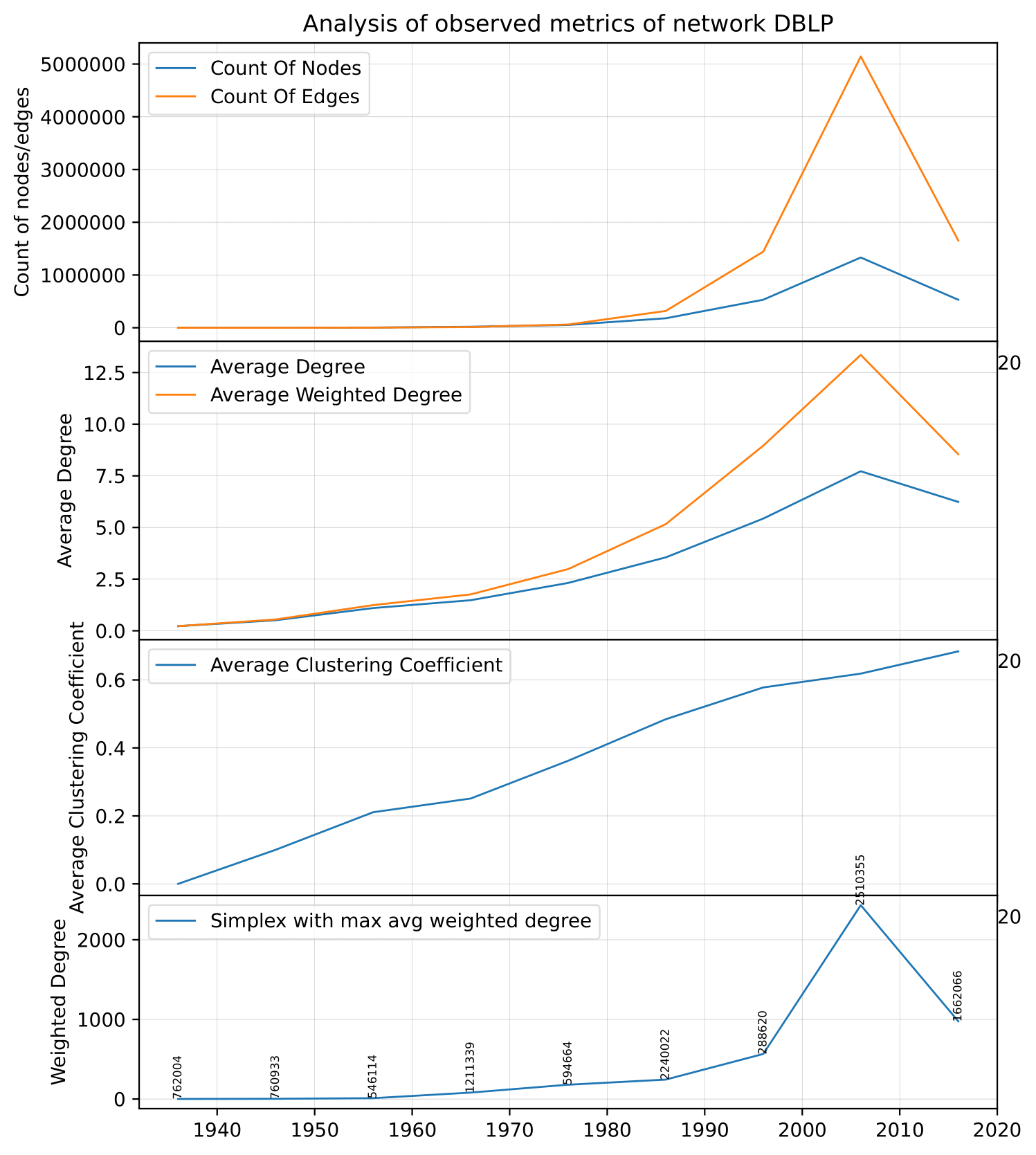
The collaboration network is typically displayed in the form of a graph, where individual authors are represented as nodes (or vertices), and the connections between them represent collaborations. Each node (author) can have connections (edges) to other authors with whom they have collaborated.

## Network attributes visualisations

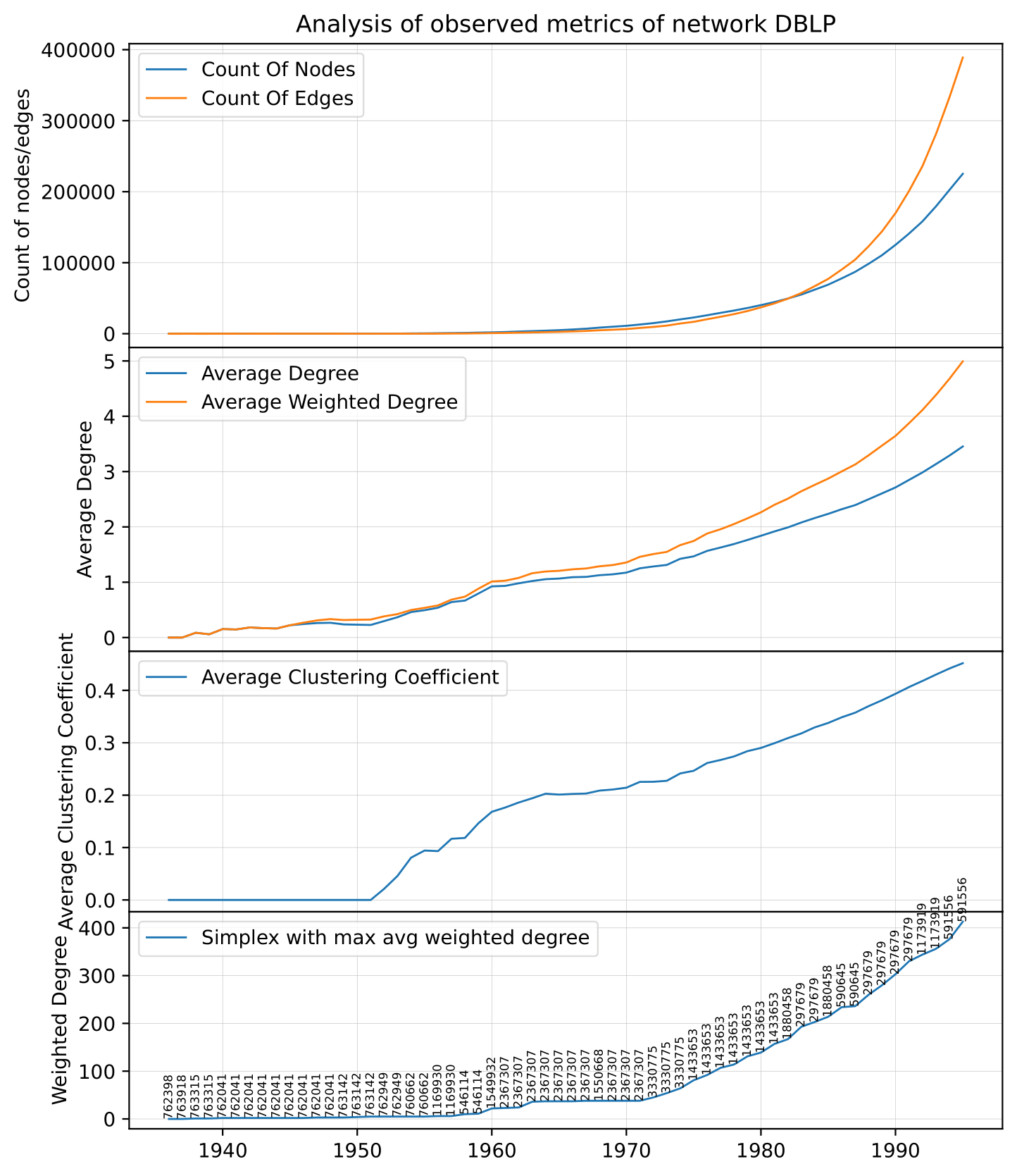
### Time step: 1 year, Not accumulated



### Time step: 10 years, Not accumulated



### Time step: 1 year, Accumulated



## Prediction of edges, DBLP

Done on prediction from year 1956 -> 1957.

Values for threshold = 0 -> Add everywhere where similarity > 0

Higher threshold means i add much more less edges.

* CN – Common Neighbors
* JC – Jaccard Coefficient
* AA - Adamic/Adar Index
* PA - Preferential Attachment
* RAI - Resource Allocation Index
* CS – Cosine Similarity or Salton Index
* SI – Sorensen Index

### Observed metrics

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Sensitivity | Recall | Specificity | Precision | Fallout | Accuracy |
| CN, 0.0 | 0.9947 | 0.9947 | 0.9998 | 0.8311 | 0.0002 | 0.9998 |
| CN, 1.0 | 0.9947 | 0.9947 | 1 | 0.9947 | 0 | 1 |
| CN, 2.0 | 0.9947 | 0.9947 | 1 | 1 | 0 | 1 |
| JC, 0.0 | 0.9947 | 0.9947 | 0.9998 | 0.8311 | 0.0002 | 0.9998 |
| JC, 0.25 | 0.9947 | 0.9947 | 0.9999 | 0.8578 | 0.0001 | 0.9999 |
| JC, 0.5 | 0.9947 | 0.9947 | 1 | 0.959 | 0 | 1 |
| JC, 0.75 | 0.9947 | 0.9947 | 1 | 0.959 | 0 | 1 |
| JC, 1.0 | 0.9947 | 0.9947 | 1 | 1 | 0 | 1 |
| AA, 0.0 | 0.9947 | 0.9947 | 0.9998 | 0.8311 | 0.0002 | 0.9998 |
| AA, 1.0 | 0.9947 | 0.9947 | 1 | 0.9492 | 0 | 1 |
| AA, 1.5 | 0.9947 | 0.9947 | 1 | 0.9947 | 0 | 1 |
| AA, 2.0 | 0.9947 | 0.9947 | 1 | 1 | 0 | 1 |
| PA, 0.0 | 0.9947 | 0.9947 | 0.8526 | 0.0053 | 0.1474 | 0.8527 |
| PA, 1.0 | 0.9947 | 0.9947 | 0.9256 | 0.0104 | 0.0744 | 0.9256 |
| PA, 5.0 | 0.9947 | 0.9947 | 0.9932 | 0.1029 | 0.0068 | 0.9932 |
| PA, 10.0 | 0.9947 | 0.9947 | 0.9995 | 0.6013 | 0.0005 | 0.9995 |
| PA, 20.0 | 0.9947 | 0.9947 | 1 | 0.9842 | 0 | 1 |
| RAI, 0.0 | 0.9947 | 0.9947 | 0.9998 | 0.8311 | 0.0002 | 0.9998 |
| RAI, 0.35 | 0.9947 | 0.9947 | 1 | 0.9492 | 0 | 1 |
| RAI, 0.5 | 0.9947 | 0.9947 | 1 | 0.9947 | 0 | 1 |
| RAI, 0.75 | 0.9947 | 0.9947 | 1 | 1 | 0 | 1 |
| CS, 0.0 | 0.9947 | 0.9947 | 0.9998 | 0.8311 | 0.0002 | 0.9998 |
| CS, 0.35 | 0.9947 | 0.9947 | 0.9998 | 0.8311 | 0.0002 | 0.9998 |
| CS, 0.5 | 0.9947 | 0.9947 | 0.9999 | 0.8821 | 0.0001 | 0.9999 |
| CS, 0.7 | 0.9947 | 0.9947 | 0.9999 | 0.9257 | 0.0001 | 0.9999 |
| SI, 0.0 | 0.9947 | 0.9947 | 0.9998 | 0.8311 | 0.0002 | 0.9998 |
| SI, 0.35 | 0.9947 | 0.9947 | 0.9998 | 0.8311 | 0.0002 | 0.9998 |
| SI, 0.5 | 0.9947 | 0.9947 | 0.9999 | 0.9257 | 0.0001 | 0.9999 |
| SI, 0.7 | 0.9947 | 0.9947 | 1 | 0.959 | 0 | 1 |

Values didnt change a lot because in this year there wasnt many common neighbors.

## Prediction of edges, Karate club

In the next table are metrics average by number of K=10 splits.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Sensitivity | Recall | Specificity | Precision | Fallout | Accuracy |
| JC, 0.45 | 0.8107 | 0.8107 | 0.9104 | 0.5936 | 0.0896 | 0.8963 |
| JC, 0.55 | 0.8077 | 0.8077 | 0.979 | 0.8584 | 0.021 | 0.9552 |
| AA, 0.35 | 0.8804 | 0.8804 | 0.6205 | 0.2784 | 0.3795 | 0.6586 |
| AA, 0.55 | 0.8412 | 0.8412 | 0.8876 | 0.5594 | 0.1124 | 0.8821 |
| AA, 0.75 | 0.8424 | 0.8424 | 0.8895 | 0.5658 | 0.1105 | 0.884 |
| CS, 0.35 | 0.874 | 0.874 | 0.7403 | 0.3536 | 0.2597 | 0.7589 |
| CS, 0.50 | 0.859 | 0.859 | 0.8977 | 0.5747 | 0.1023 | 0.8922 |

# Vzít všechny hrany a rozdělit si je na 10 částí. K fold cross validace. 9 nechám, jednu vynechám. A znovu.

# vynechat 10 % hran a označit si je, že jsou vynechané.

# Spustit algoritmus na všechny vrcholy a vytvořit si tabulku s TP, TN, FP, FN.

# Spustit algoritmus na vrcholy, u kterých se týkalo odstranění. # NN

#Nebo si vezmu první snímek a kouknu se do druhého snímku.