



On the Relation of Edit Behavior, Link Structure, and Article Quality on Wikipedia

Thorsten Rupprechter, Tiago Santos, Denis Helic

Graz University of Technology

rupprechter@tugraz.at, tsantos@iicm.edu, dhelic@tugraz.at

Motivation

Understanding the relation between editing behavior, link structure, and article quality may facilitate editing processes and article quality on Wikipedia, especially considering editor conflicts. Hence, we ask the following research questions:

- How can we **characterize editing** behavior on Wikipedia?
- Is there a **relation** between **editing behavior**, **article quality**, and **wikilink network topology** on Wikipedia?
- If such a relation exists, **how strong** is it?

Edit Behavior Modeling and Wikilink Network Analysis

Article Quality Category Dataset

- **13,045 labeled article histories** from Wikipedia categories Featured, Good, A, B, and C as well as edit war articles
- Closely analyzed **4,800 article histories** (800 per category)

Edit Action Labels

Content	Format	WikiContext
Fact-Update	Refactoring	Vandalism
Verifiability	Copy-Editing	Counter-Vandalism
Simplification	Wikification	Processing Tags
Clarification	Link-Disambiguation	Others
Elaboration		
Point of View		

Edit Label Metrics

- **Relative Edit Label Frequency** of labels in article histories
- **Label Transition Probability** of first-order Markov chains, resulting from label transitions in article edit histories

Network of Wikilinks

- Utilize **WikiLinkGraphs** framework to build graph
- **Articles** are **nodes**, **wikilinks** between articles are **edges**
- Compute corresponding parsimonious **network metrics**, e.g., clustering coefficient and PageRank

Results

Our analysis unveils that **controversial and edit war articles (CW)** differ considerably from others for almost all metrics. Additionally, slight trends are detectable for higher-quality articles such as Featured (FA) or A. Our results may pave the way for usage of link metrics as controversy indicators, further development of Wikipedia content-assessment tools, or editor role identification.

Relative Label Frequency and Transition Probability

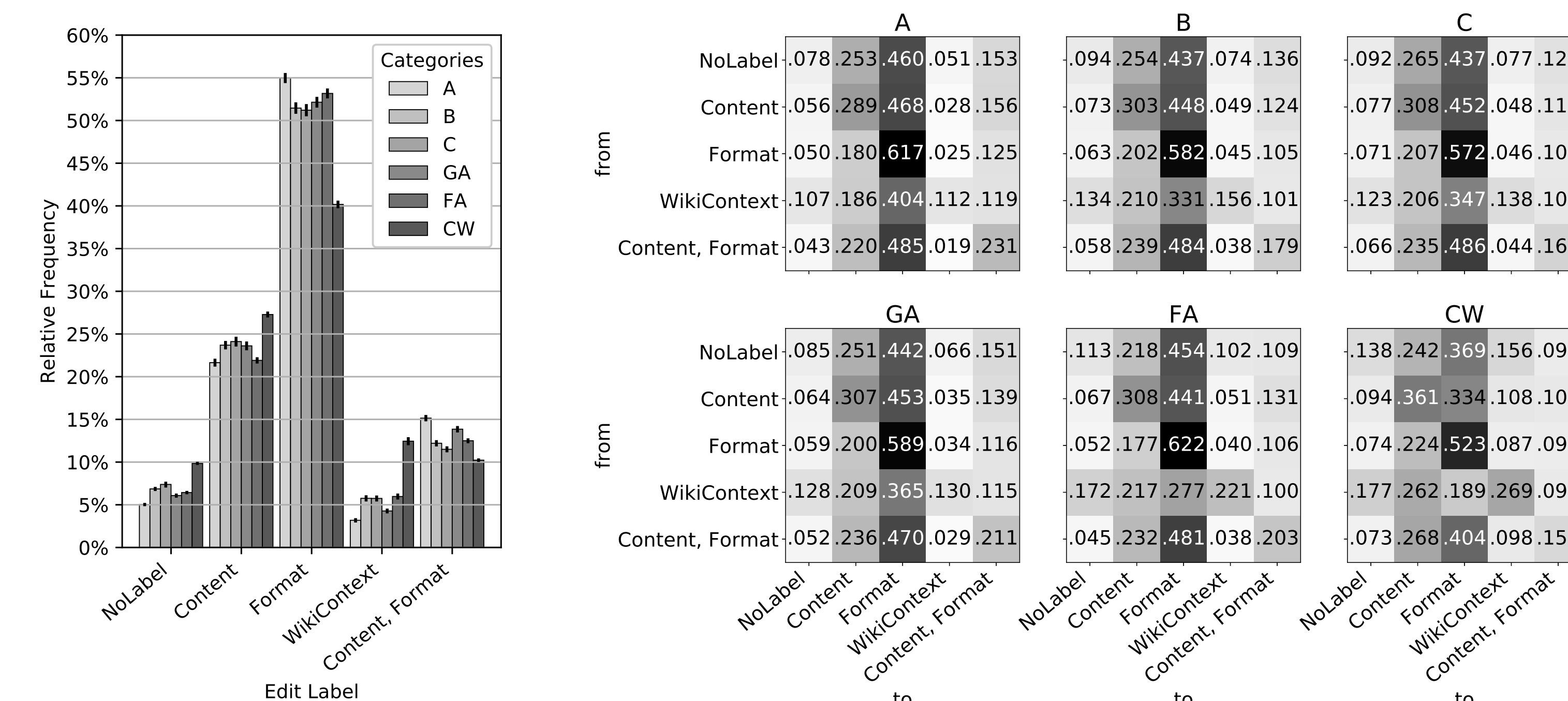


Figure 1: Relative edit label frequencies (left) and label transition probabilities (right).

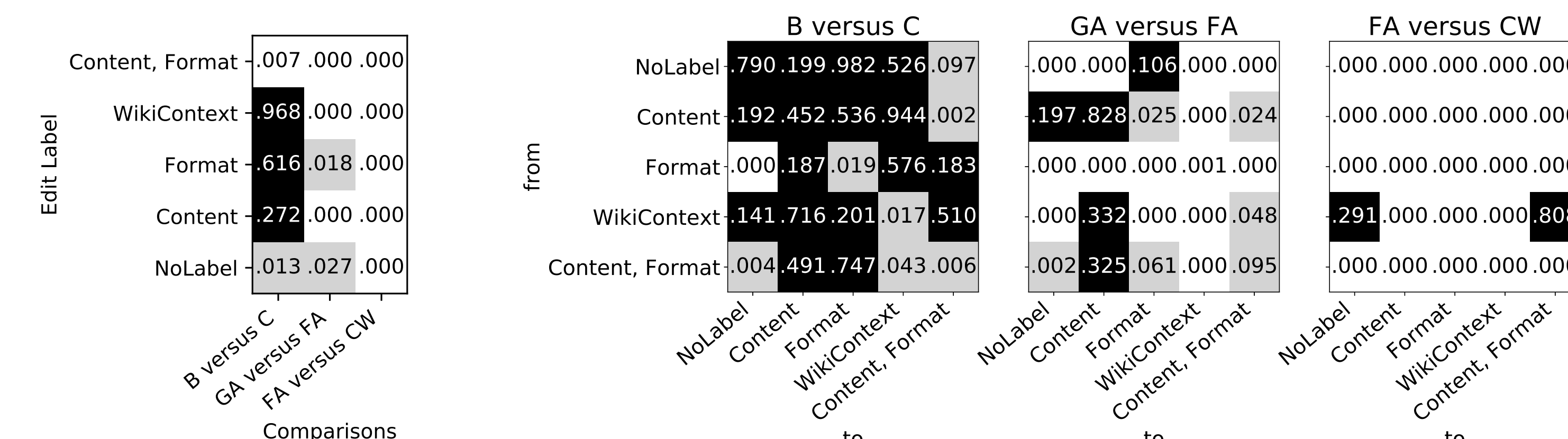


Figure 2: P-values of permutation tests for relative label frequencies (left) and label transition probabilities (right) for selected article categories.

Network of Wikilinks

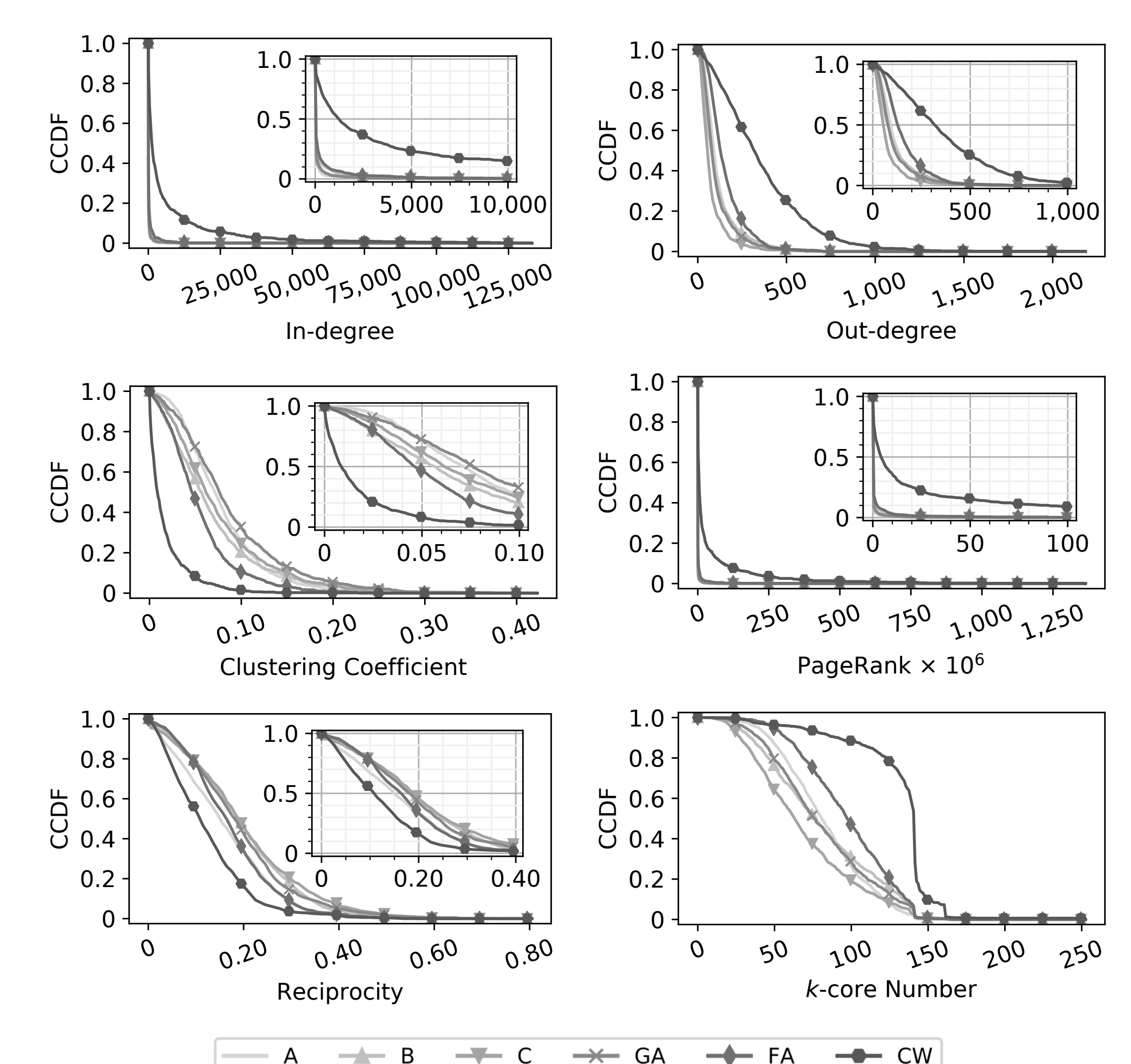


Figure 3: Complementary cumulative distribution functions (CCDF) of network metrics across multiple quality categories, with insets highlighting relevant chart areas.

Tiago Santos is a recipient of a DOC Fellowship of the Austrian Academy of Sciences at the Institute of Interactive Systems and Data Science, TU Graz.