PhD candidate - Justin Wang Ngai Yeung: Developing Criminal Networks with Inadequate Data

Description

Justin Wang Ngai Yeung, from Network Science Institute at Northeastern University in London, highlighted the challenges of social network analysis during his appearance in the Data Skeptic" podcast. The conversation emphasized the limitations of techniques such as weighted graphs when working with incomplete and unreliable data from law enforcement agencies. Yeung underscored that data collection in this context is highly nuanced, prone to human error, and ultimately reflects the familiar phrase "garbage in, garbage out".

Something I learned

Most interestingly, Yeung claims that he needs to understand 50-60% of a criminal network to predict an organization's taxonomy and hierarchy. Yet, in most cases, there is limited data, often representing <50% of the criminal network, to describe each node's (person's) relationships (interaction) with other nodes (persons). There is a paradox of having an abundance of data, much of which is "garbage" or incomplete, making it extremely difficult to identify significant patterns and impactful relationships with a high degree of confidence.

Connection to INFO 526

Throughout the INFO 526 course, we've worked with large datasets, applying data wrangling techniques such as reformatting, manipulation, transformation, and summarization. This story highlights the importance of these skills and gives me a new perspective on the need for algorithmic solutions, particularly in social network analysis, to gain meaningful insights from complex, relational data. These methods are paramount for constructing informative network graphs and their respective relationships, enabling a deeper understanding of hidden structures and hierarchies within a network.

Source: Data Skeptic Podcast. *Criminal Networks*