Visualizing Rumor Propagation in the PHEME Dataset

By Chuhan Ni

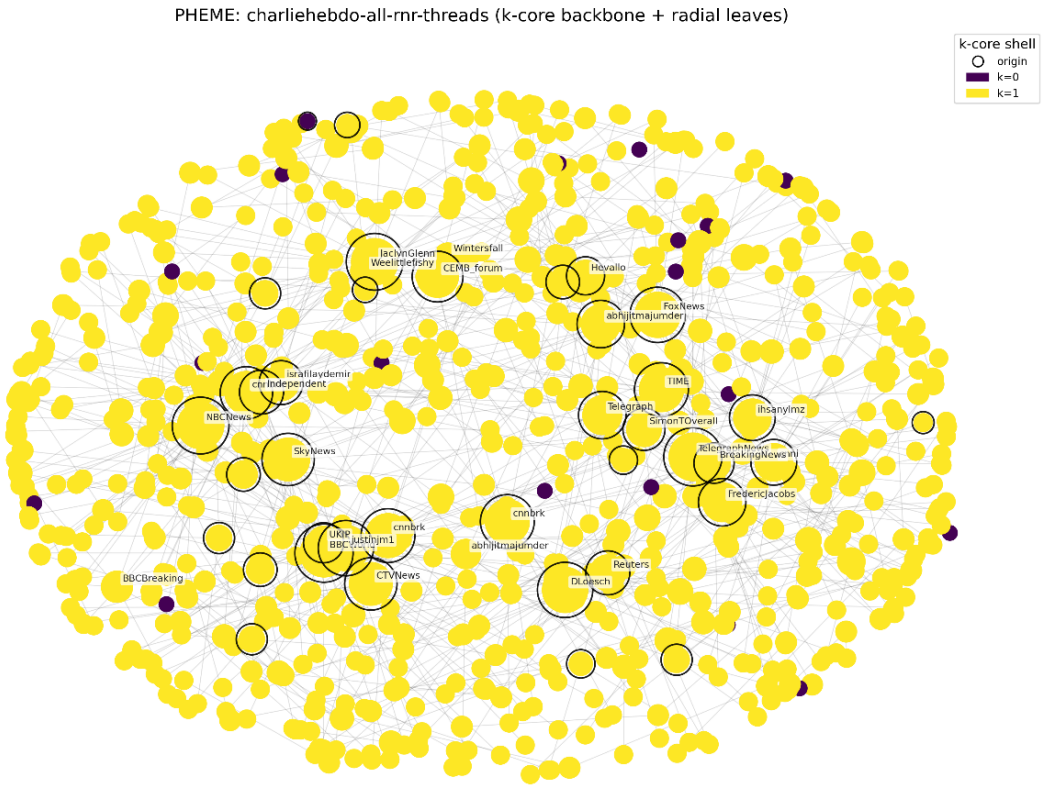

Figure 1. Rumor Propagation Network

Figure 1 presents an integrated visualization of rumor propagation during the Charlie Hebdo event. Each node in the network represents an individual Twitter user who participated in the discussion, while edges indicate reply or retweet relationships that form the overall information flow. The size of each node corresponds to its degree, meaning that larger nodes represent accounts with greater influence or engagement within the conversation. Node colors reflect each user's k-core shell value: bright yellow denotes users located in the densely interconnected inner core, whereas darker purple marks peripheral users who are only loosely connected to the network.

Origin tweets — the first posts that initiated each discussion thread — are highlighted with black outlines. These origins appear dispersed across the network, indicating that multiple independent tweets triggered parallel conversations rather than a single viral cascade. The layout combines spring-based positioning for the k-core backbone with radially placed peripheral users, producing a clear structural separation between core broadcasters and reactive audiences. Overall, the visualization reveals a strong core–periphery pattern typical of large-scale rumor diffusion. Prominent media outlets such as BBCBreaking, CNN, Reuters, SkyNews, and NBCNews dominate the central region, functioning as primary information hubs. Surrounding them, a large number of peripheral users contribute minimal structural influence, largely resharing core content rather than generating new conversational branches. This structure suggests that information flow in crisis events is driven not by mass participation but by a relatively small cluster of interconnected, authoritative accounts whose messages propagate outward to a wide audience.

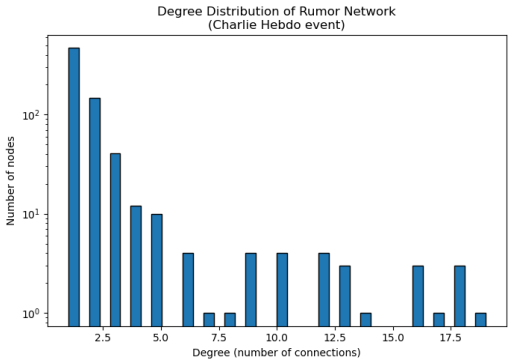

Figure 2. Degree Distribution

Figure 2 illustrates the temporal dynamics of rumor-related tweets over the duration of the Charlie Hebdo event. The majority of activity occurs within the first ten hours, indicating a rapid initial surge of attention typical of crisis-driven conversations. A secondary wave of activity emerges between roughly twenty and fifty hours, suggesting that additional developments or renewed public interest contributed to revived engagement. Beyond this period, tweet activity decreases sharply, demonstrating a natural decay in attention as the event loses immediacy. A small number of late tweets appear well after one hundred hours, reflecting residual discussions or delayed reactions by peripheral users. Overall, the temporal distribution reveals that rumor diffusion follows a burst-and-decay pattern, characterized by intense early activity followed by fragmentation and decline—a pattern commonly observed in breaking news cycles on social media.

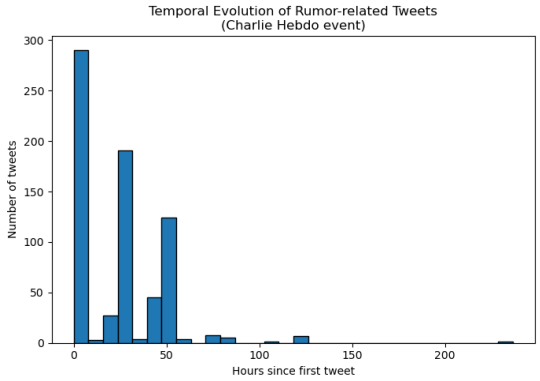

Figure 3. Temporal Evolution

Figure 3 displays the degree distribution of the rumor network on a logarithmic scale. The visualization shows that most users possess very few connections, typically one to three, forming a broad and sparsely linked periphery. In contrast, a small number of nodes reach substantially higher degrees, confirming the presence of a heavy-tailed or scale-free–like distribution. Such a structure indicates that the network is organized around a small set of highly connected hubs whose removal would significantly disrupt communication, while the removal of random peripheral users would have little structural impact. This suggests that rumor propagation relies heavily on a limited number of influential accounts that act as central conduits for information flow. The pattern is consistent with social media ecosystems, where visibility and impact are unevenly distributed.

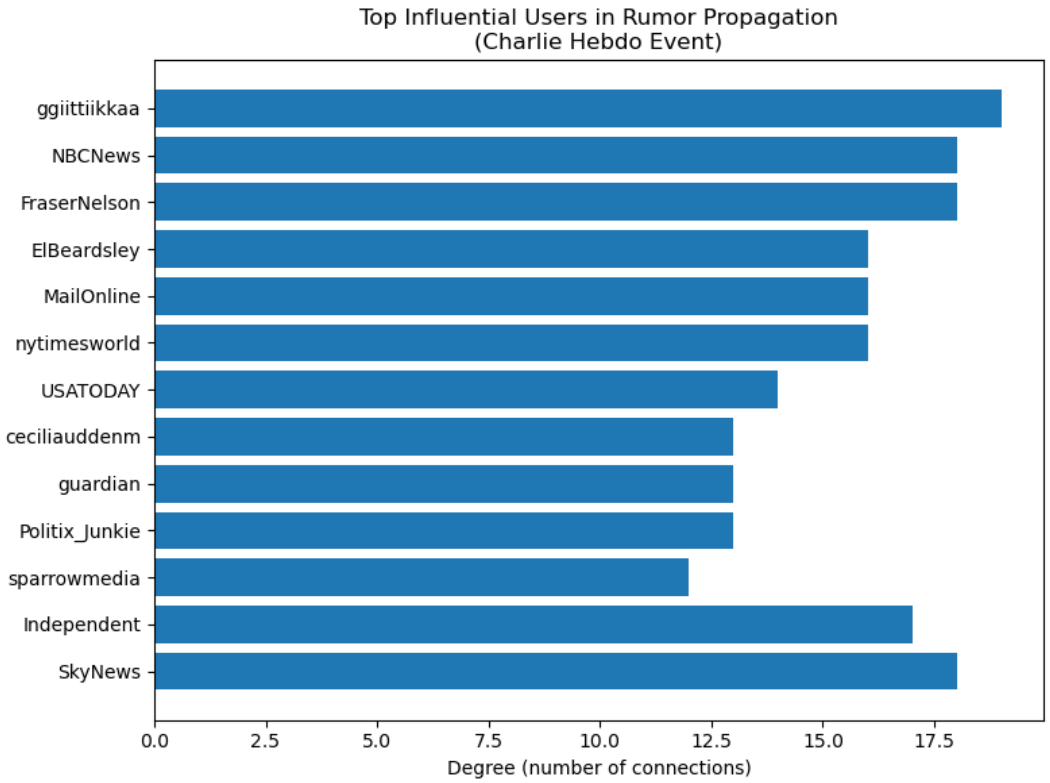

Figure 4. Top Influential Users

Figure 4 highlights the most influential users within the rumor network by comparing their number of connections to other participants. The accounts with the highest degrees tend to be professional news organizations, including SkyNews, NBCNews, MailOnline, and USATODAY. Their prominence underscores the central role that established media outlets play in shaping early narratives and amplifying event-related information. A small number of individual users, such as *ggijittikkaa*, also appear among the top-ranked accounts, demonstrating that influence is not restricted solely to institutional profiles. The distribution of bar lengths further suggests a heavy-tailed structure: while a few users accumulate a large number of interactions, the vast majority of participants maintain only a few connections. This imbalance aligns with broader patterns observed in information diffusion, where a small elite of broadcasters drives most of the visibility and circulation of content.

This visualization is based on the PHEME dataset of rumours and non-rumours on Twitter, focusing on the Charlie Hebdo event. I first extracted all annotated threads for this event and parsed the tweet JSON files to obtain user IDs, timestamps, and reply relationships. These threads were merged into a single conversation graph where nodes represent users and edges represent reply or retweet interactions between them. Using Python, I computed node degrees and k-core indices with NetworkX, applied a spring layout to the k-core backbone, and positioned peripheral nodes radially for clarity. Additional summary visualizations, including the degree distribution, temporal evolution of tweet volume, and top influential users, were then generated using Matplotlib.

Data source: https://figshare.com/articles/dataset/PHEME_dataset_of_rumours_and_non-rumours/6392078

Github link: https://github.com/jollyjellyjewel/Information-Visualization-Report