Social Media and Network Sciences

# Twitter Retweet Network Analysis

DSM160-2024

# Abstract

X (formally Twitter), is a massively popular social network, formally categorized as a microblogging network. A network where followers follow influencers, and opinions (Tweets) by influencers are broadcasted with a specific Hash-tag/s (#ChatGPT for example), these tags buildup a social circle encompassing the communications in the form of Tweets, Retweets, Likes etc. We will be analyzing a collection of Retweet data gathered from Twitter’s API access for 6 tags, #chatgpt, #bitcoin, #BoredApeYatchClub, #Eth, #uranium. The report will explore each networks structure using Network science techniques and a comparative analysis will highlight the differences between them. For this purpose, a Retweet network will be constructed with each Original tweeter and re-tweeters assigned as Nodes and their link between them will specify the flow of information.

# Keywords

Network Science, Twitter (now X), Retweet analysis, Crypto tweet network

# Data

The data being analyzed is a static collection of these networks, hence the analysis will also focus on the state of things at the specific time and duration. There is incentive to have a larger time frame for analysis in order to extract a deeper and accurate understanding of the network, below we define the data in terms of its 4 Vs.

### Volume

Each retweet data contains in addition to the tweet text, a large amount of meta data about the tweeter, and in the case of it being a retweet, it also contains the metadata for the original user. From the provided extract of data, an average Tweet is 6KB in size, one day data amounts to 25MB.

### Velocity

Tweet rate for the networks is as below, #Etherium and #ChatGPT are similar and the fastest in terms of new data generation. With 6KB of new data generated ever ~30sec.

|  |  |  |  |
| --- | --- | --- | --- |
| **Network** | **Rate of Tweet** **(H:MM:SS)** | **Measurement Period** | **Date** |
| #ChatGPT | 0:00:29.861600 | 1 day, 17:28:28 | 2023-02-06 |
| #Bitcoin | 0:00:16.338431 | 22:51:04 | 2023-02-07 |
| #boredApeYachtClub | 0:02:42.061663 | 9 days, 12:30:25 | 2023-02-27 |
| #energy | 0:00:31.479800 | 1 day, 19:43:19 | 2023-02-06 |
| #eth | 0:00:28.380324 | 1 day, 15:25:30 | 2023-02-07 |
| #uranium | 0:01:43.696600 | 6 days, 0:01:23 | 2023-02-06 |

### Variety

The retweet data is contained in a json file, for the purpose of this analysis we are limiting ourselves to the text-based metadata of each retweet. However, the file also contains links to videos, images and weblinks.

### Veracity

In a social media setup, there is a good chance of data misrepresentation or irrelevancy to the topic, other factors such as bots and spams also contribute to the challenges faced during analysis. Which is why it is of benefit to identify these players.

### Legislation and Bot/Spam consideration

## General Data Protection Regulation (GDPR) imposes strict rules on personal data handling (https://gdpr-info.eu­­), it operates in the European union. It defines a legislative ground by defining the principles, rights, restrictions such as, Lawful processing, Consent, code of conduct, transfer of data and penalties.

## Other sources to keep in mind are X’s own term of service and developer agreement, where the sharing restriction of API accessed data and its utilization for surveillance.

Misinformation and Bot/Spam content ….

# Processing the Data

The json files are loaded into python using the json library, for each re-tweet in the network, the information on the original tweeter and the re-tweeter are captured along with their follower counts. These users are designated as nodes, each re-tweeter is linked to its original tweeter via the edges. Larger the interaction between two such nodes, greater will be the weight of the edge. The nodes themselves are weighted by the number of their followers and also attributed by the number of tweets/re-tweets sent by them. Networkx library is used for the creation of the networks. Pyvis library is used for creating the visualization for each of the networks.

The resulting networks are directed and weighted as per the above-mentioned attributes. Within each network, there are many weakly connected components, visualized as a star like structure, where each compartment acts as its own re-tweet network centered around an original tweeter (influencer). The directivity of links represents the flow of information from the tweeter to the re-tweeter.

### Network of Chatgpt

Network of bitcoin

Network of etherium

Network of B

Network of energy

Network of uranium

### Multipath propagation

An electromagnetic wave when traveling through the environment may interact with objects either by diffracting, reflection, absorption, or scattering, these give rise to secondary waves that are received at the receiver with varying propagation delays. The superposition of these waves at the receiver with their angular and phase shifts can be constructive or destructive at the receiver, resulting in the Fast fading phenomenon.

In NLOS conditions, signal experiencing rich scattering, with a high number of waves superimposed the amplitudes of the phasors become Rayleigh distributed, the fading in such conditions is characterized as Rayleigh fading. If however, one wave is much stronger than the others, then the distribution can be characterized as Ricean (resulting in Ricean fading).

### Propagation Channel

The directional propagation channel can be described by its propagation path, which in turn is a formulation of its properties as follows. 1. Time delay (τn ), 2. The direction of departure (azimuth angle ϕt;n zenith elevation angle θt;n) 3. The direction of arrival (azimuth angle ϕr;n zenith elevation angle θr;n), the relation between the transmitted Et;n and received electric field Er;n.

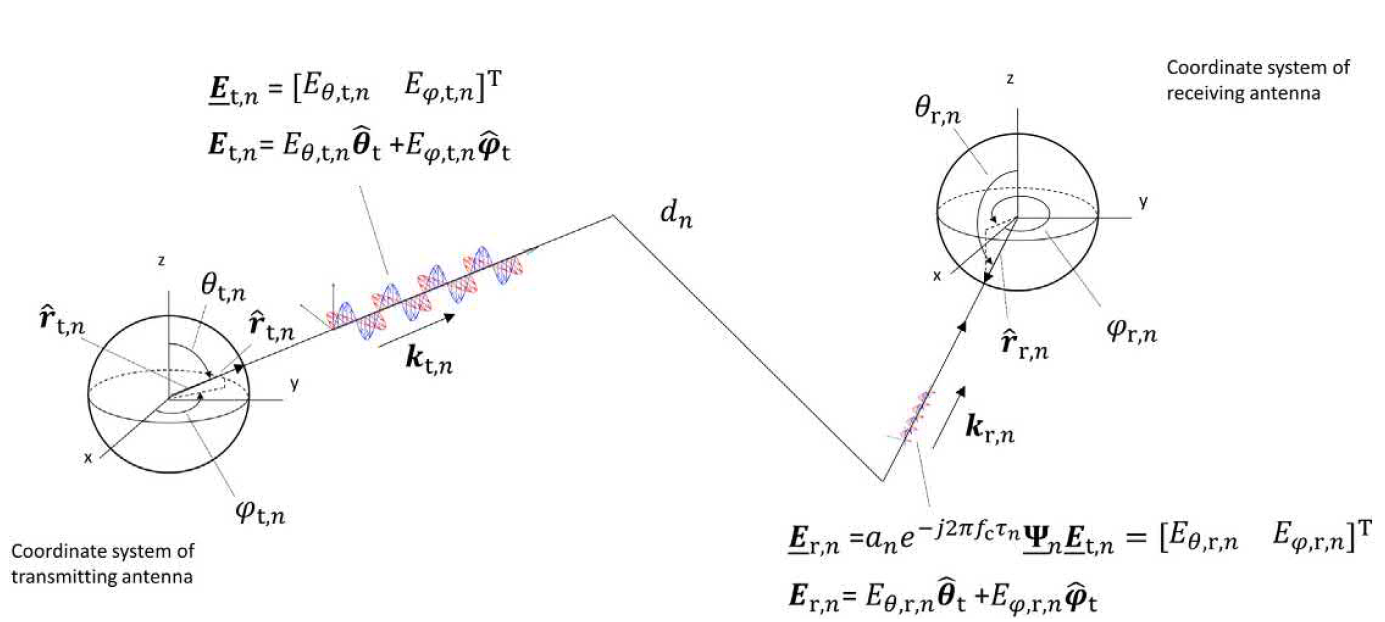


Figure 1 Properties of the Propagation Channel [12]

The end to end tranmission can then be denoted by defined by below equation; where Y(f) is the received signal, H(f) is defined as the Channel transfer function and X(f) is the transmitted signal.

### Massive MIMO

Massive MIMO systems use a large number of antennas to transmit and receive multiple data streams, which increases the capacity and reliability of the antenna system, Massive MIMO is generally achieved by Active Antenna System (AAS) which integrates the antennas and transceivers into a single unit. mMIMO can be implemented for both Frequency division duplex (FDD) and Time division duplex (TDD) modes, however, most of the work has been focused on utilizing the reciprocity property of TDD.

#### Beamforming

mMIMO allows for an increase in gain of transmission towards a particular direction thereby reducing it in others. Multiple beams are possible in a single Frequency and Time. The beam creation is done via an array of transceivers, which are positioned in Horizontal and vertical columns to create a 2D array.

Tranmission matrix H contains the impulse response between the tranmitter and receovier in a mMIMO transmission.

A diagram of a building with a tower and a tower with a tower and a tower with a tower and a tower with a tower and a tower with a tower and a tower with a tower and

Description automatically generated

Figure 2 Evalution from Passive to 2D Active antenna system [13]

Considering a Path from Base Station (BS) towards User equipment (UE), the beamforming in AAS is done by a process called precoding, where beam weights are assigned for the creation of the beam. The receiver which is usually a single or dual antenna system implements spatial combining.

#### Channel Response and CSI feedback

As discussed earlier the transmission medium is fast changing due to myriads of reasons, beamforming weights need to be optimized or selected correctly in order to fully utilize the gains. Pilots are sent from the Transmitter identifying the beams these pilots help in channel estimation and decoding of the user data. The receiver after Channel estimation then has to feedback the Channel State Information (CSI) to the transmitter in order to update the beamforming weights.

A diagram of a channel diagram

Description automatically generated

Figure 3 CSI acquisition from closed-loop CSI feedback [12]

A white background with black and white lines

Description automatically generated

Figure 4 Sparsity of the Channel Matrix in mMIMO - L=128 [5]

# AI in Wireless networks

The use of AI in the telecommunications ecosystem, particularly in wireless systems, can be observed through the interaction between two entities: the Base Station and the User Equipment (UE). Both employ independent Deep learning systems for enhancing their functionality.

### AI at the Base Station

#### Cognitive Optimization (Anomaly Detection, Self Healing)

A live radio network undergoes constant monitoring and optimization in order to cater for changing and growing user behavior, and environmental changes. Several key KPIs are monitored to evaluate the quality, availability, and reliability of each cell’s performance in the network. Actions are taken per use case scenario, for example, a new device introduced in the market started to drop connection due to a missing device capability, or a coverage hole is detected resulting in loss of service for consumers in a particular location due to a physical blocking by a newly erected building, etc. Such cases require parametric tuning which could be localized to a single cell or generalized to the complete network. A part of cognitive optimization, Anomaly detection enables zero-touch operation. A properly trained AI system can detect and self-heal an anomaly and monitor for required minimization of error. These systems can result in reduced man hours, 24-hour operation, and fast response times while avoiding human error.

#### Network Planning

The expansion of a network requires multiple inputs upon which a decision to invest can be based. An AI model for coverage and capacity analysis utilizing big data capabilities can be utilized, which takes as input Crowdsourced data for sentiment analysis, population density, area classification, and proactive congestion detection [17] to make autonomous changes to antenna tilts, bandwidth allocation, Load balancing, Access control, etc.

#### Interference mitigation

Wireless networks are prone to internal and external interference, and both Uplink and Downlink channels have their set of challenges. AI models trained to detect and take corrective actions are being utilized in current 5G networks.

#### Power Optimization

Base Station (BS) power saving can be achieved by intelligent sleeping off some cells. Energy savings can even be achieved at micro levels by turning off unused transmission slots. Intelligent forecasting is required to take such actions in order to avoid coverage impact.

### AI at the UE End and Two-sided models

Currently, there is active work in 3GPP release 18, for defining a framework for AI at the UE and networks joint models [18], some of the use cases are listed below.

#### Chanel Estimation and CSI Feedback

Prediction of channel coding [Mahmoud, 2021], Channel estimation for massive-MIMO [1],[2],[3],[4],[5],[6],[7],[8],[9],[10], CSI compression

#### Beam management

Spatial and Temporal - domain beam prediction for a set of beams from the measurement results of another beam set. [18](3GPP 38.384)

#### Position Accuracy enchancments

AI/ML-based UE positioning, i.e fingerprinting (Fingerprinting in the context of positioning refers to a technique that uses signal strength data from beacons or sensors to determine a user’s location indoors) [18](3GPP 38.384)

# Deep Learning for Massive MIMO

Deep learning solutions have been proposed as a redesign for several aspects of wireless transmission, especially for mMIMO, during the process of encoding, decoding, beam formation, beam detection, antenna selection, and channel estimation.

large number of antennas in an mMIMO system increases the spatial complexity of the system making it less efficient using available algorithms. Deep learning results have shown similar or improved performance in many of the aspects of wireless transmission.

Here some recent work has been reviewed in the specific domain of channel estimation in mMIMO systems

## Channel estimation for CSI feedback

### Encoder and Decoder architecture for CSI feedback in FD mMIMO systems

In [C. -K. Wen, 2018] tackle the issue of correct CSI feedback report for FD mMIMO, CSI feedback (Pilot) overheads are considered unwanted and the author argues that the alternate method of using sparsity of the channel is never guaranteed and hence CS (Compression sensing) techniques are unreliable. An Encoder at the UE end learns the transformation from the original channel matrices to compress codewords through training data. This is feedback to the BS which uses a Decoder to perform inverse transformation to recover the CSI.

A diagram of a diagram

Description automatically generated

Figure 5 (a) Pseudo color for the strength of H, (b) Encoder-Decoder architecture [6]

### Pilot detection with Compressed sensing and Two stage channel estimation

In [C. -J. Chun, 2019] author discusses the overheads and computation cost associated with typical Pilots in mMIMO systems since typically the Pilot length (Ls) is equal to or larger than the number of transmit antennas (Nt), without which the linear mean square error (LMMSE) performance is substantially degraded. LMMSE is commonly used for channel estimation, it is computationally costly, and complexity increase with longer Ls. Additionally, in mMIMO systems with very high antenna numbers (such as mili meter wave), it will not be feasible since Ls should not be larger than the channel coherence interval ([D. Neumann, 2015] proposed alternatives such as a Compression-based linear minimum mean square error (CLMMSE)).

In the condition where Ls < Nt, [C. -J. Chun, 2019] proposes a two-stage channel estimator, the first stage of the transmitted (with the pilot), takes as input the received channel (h) and weight matrix of the pilot. Offline training of the TNN is done by feeding the weight matrix and resulting pilots. The resulting pilot is fed to a DNN which takes the received signal (ys) including the noise to estimate the channel. This DNN acts as a non-linear channel estimator.

In the subsequent slots, the second DNN is fed with estimated data X(k) using the traditional LMMSE detector along with the received signal yx resulting in a predicted channel estimation h(x).

### Using an Offline-Online two-step training mechanism for CSI detection

In [L. Ge, 2019] the authors focus on real-world features affecting the channel response such as Frequency, band, location, time, temperature, humidity, and weather. The spatial-temporal relationship of CSI to design a learning framework of a convolution neural network and a long short term with memory (LSTM) network. An Offline-Online system two-step training mechanism is utilized.

### Channel estimation for Millimeter-Wave system using a 3D CNN + RNN

In [W. Luo, 2021] channel estimation for a 3D Massive MIMO in a millimeter wave system is proposed. High-speed mobility is taken into consideration when designing the 3D CNN. Transmit Antennas NT denoting the spatial aspect, Nc referring to the frequency (number of OFDM carriers), and T as the time when sampling the Doppler frequency shift, represent a three-dimensional matrix (NT,Nc, T) which are fed to a 3D CNN and trained offline.

# A diagram of a complex flowchart Description automatically generated

Figure 6 The Network structure [7]

### RIDNet-based DeNoising channel estimation

In [R. He and W. Zhou, 2022] The Author builds up on earlier denoising networks for CSI detection such as [M. Soltani, 2019] Denoising convolution neural network (DcNN), [Y. Jin, 2019] Flexible denoising convolution neural network (FFDNet), and [Y. Jin, 2020] Convolution Blind denoising Network (CBDNet) argues that all these techniques required manual intervention or were not efficient. Based on this argument a RIDNet-based channel estimation architecture is introduced, additionally, the Multi-User MIMO scenario is considered for the first time. The channel model used in this paper utilizes the Angle of Arrival of the users to construct a sparse matrix

A diagram of a machine learning

Description automatically generated

Figure 7 Architecture of the RIDNet-based channel estimator [5]

# Conclusion and Discussion

mMIMO is a pivotal technology enabling the realization of spectral efficiency promised in 5G networks. The papers discussed above utilized Offline trained Deep Learning modules to generate inferences regarding the received channel, and varying level of complexities were considered. Earlier papers focused on a general AI/ML model which applied CSI compression and decompression at the receiver, with an argument regarding higher overhead transmission and unreliability of existing algorithms. building on this later papers focused on amplifying the accuracy in real-world scenarios by considering the high-moving scenario, Rician vs Rayleigh interference considerations, and real-world features (frequency, timer, location etc). Denoising models provide higher accuracy by removing the noise in the radio propagation. With Release 18 3GPP has embraced the AI/ML direction both in UE-side (inference is performed at the UE) and Network-side models. Top uses cases include the CSI detection at the UE side and CSI compression and transmission at the network are being considered for testing and standardization.

# References

[1] C. -J. Chun, J. -M. Kang and I. -M. Kim, "Deep Learning-Based Channel Estimation for Massive MIMO Systems," in IEEE Wireless Communications Letters, vol. 8, no. 4, pp. 1228-1231, Aug. 2019, doi: 10.1109/LWC.2019.2912378.

[2] D. Neumann, M. Joham, L. Weiland and W. Utschick, "Low-Complexity Computation of MMSE Channel Estimates in Massive MIMO," WSA 2015; 19th International ITG Workshop on Smart Antennas, Ilmenau, Germany, 2015, pp. 1-6.

[3] L. Ge, Y. Zhang, G. Chen and J. Tong, "Compression-Based LMMSE Channel Estimation With Adaptive Sparsity for Massive MIMO in 5G Systems," in IEEE Systems Journal, vol. 13, no. 4, pp. 3847-3857, Dec. 2019, doi: 10.1109/JSYST.2019.2897862.

[4] C. Luo, J. Ji, Q. Wang, X. Chen and P. Li, "Channel State Information Prediction for 5G Wireless Communications: A Deep Learning Approach," in IEEE Transactions on Network Science and Engineering, vol. 7, no. 1, pp. 227-236, 1 Jan.-March 2020, doi: 10.1109/TNSE.2018.2848960.

[5] R. He and W. Zhou, "Channel Estimation for Massive MU-MIMO Systems with Real Image Denoising Network," 2022 7th International Conference on Computer and Communication Systems (ICCCS), Wuhan, China, 2022, pp. 513-518, doi: 10.1109/ICCCS55155.2022.9845979.

[6] C. -K. Wen, W. -T. Shih and S. Jin, "Deep Learning for Massive MIMO CSI Feedback," in IEEE Wireless Communications Letters, vol. 7, no. 5, pp. 748-751, Oct. 2018, doi: 10.1109/LWC.2018.2818160.

[7] W. Luo, W. Ji, Y. Song and B. Zheng, "Deep Learning-based Channel Estimation Approach for 3D Massive MIMO Millimeter-Wave System in Time-Varying Environments," 2021 IEEE/CIC International Conference on Communications in China (ICCC), Xiamen, China, 2021, pp. 529-533, doi: 10.1109/ICCC52777.2021.9580417.

[8] M. Soltani, V. Pourahmadi, A. Mirzaei and H. Sheikhzadeh, “Deep Learning-Based Channel Estimation,” in IEEE Communications Letters, vol. 23, no. 4, pp. 652-655, Apr. 2019.

[9] Y. Jin, J. Zhang, S. Jin and B. Ai, “Channel Estimation for Cell-Free mmWave Massive MIMO Through Deep Learning,” in IEEE Transactions on Vehicular Technology, vol. 68, no. 10, pp. 10325-10329, Oct. 2019.

[10] Y. Jin, J. Zhang, B. Ai and X. Zhang, “Channel Estimation for mmWave Massive MIMO With Convolutional Blind Denoising Network,” in IEEE Communications Letters, vol. 24, no. 1, pp. 95-98, Jan. 2020.

[11] [Cellular-Communications-in-a-5G-Era-InDesign.pdf (5gamericas.org)](https://www.5gamericas.org/wp-content/uploads/2022/01/Cellular-Communications-in-a-5G-Era-InDesign.pdf)

[12] <https://www.researchgate.net/publication/342590194_Advanced_Antenna_Systems_for_5G_Network_Deployments_-_Bridging_the_Gap_Between_Theory_and_Practice>

[13] [Massive MIMO for New Radio | Samsung Business Global Networks](https://www.samsung.com/global/business/networks/insights/white-papers/1208-massive-mimo-for-new-radio/)

[14] H. H. H. Mahmoud and T. Ismail, "A Review of Machine learning Use-Cases in Telecommunication Industry in the 5G Era," 2020 16th International Computer Engineering Conference (ICENCO), Cairo, Egypt, 2020, pp. 159-163, doi: 10.1109/ICENCO49778.2020.9357376.

[15] Mahmoud, Haitham & Amer, Amira & Ismail, Tawfik. (2021). 6G: A Comprehensive Survey on Technologies, Applications, Challenges, and Research Problems. Transactions on Emerging Telecommunications Technologies. 1-14. 10.1002/ett.4233.

[16] <https://www.ericsson.com/en/blog/2021/9/why-intelligent-ran-automation-is-going-to-rock-your-world>

[17] <https://www.ericsson.com/en/blog/2020/12/ai-and-network-capacity-planning-decisions>

[18] [Specification # 38.843 (3gpp.org)](https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3983)

[19] [Whitepaper-The-future-of-AI-is-hybrid-Part-1-Unlocking-the-generative-AI-future-with-on-device-and-hybrid-AI.pdf (qualcomm.com)](https://www.qualcomm.com/content/dam/qcomm-martech/dm-assets/documents/Whitepaper-The-future-of-AI-is-hybrid-Part-1-Unlocking-the-generative-AI-future-with-on-device-and-hybrid-AI.pdf)