SAND2022-12345C

Sandia National Laboratories

Exceptional service in the national interest

A Simulation Test Bed for Evaluating Data Analytic Predictors of Disinformation Flow

Kiran Lakkaraju, 8716

September, 20th, 2022

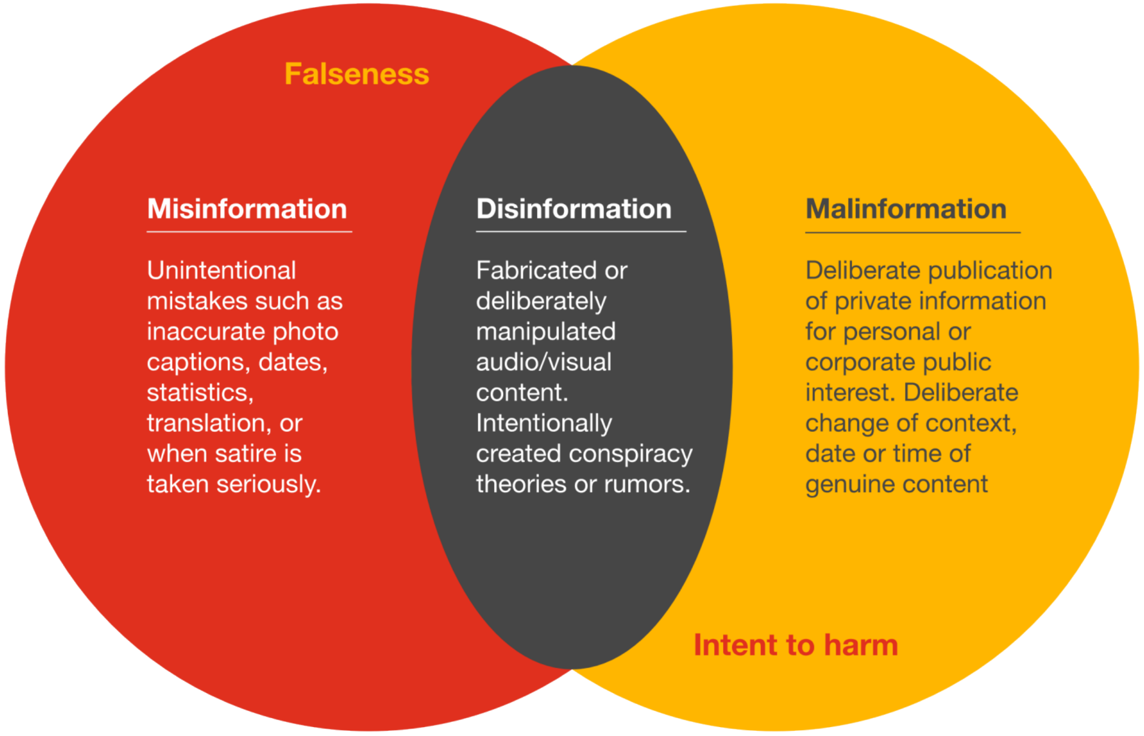
SBP-BRIMS 2022

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

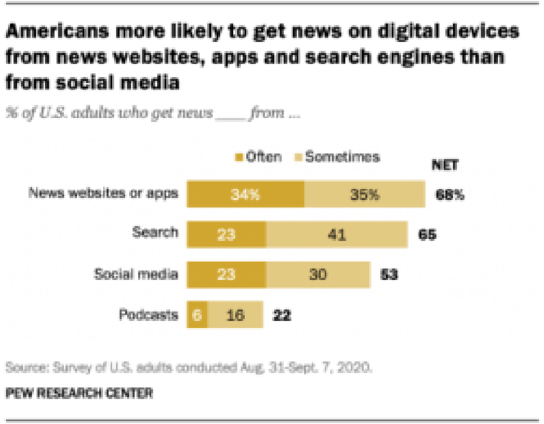
Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Disinformation is being used by many nation-states

- Disinformation is false information intentionally used for harm.
  - Nation-State and non-state actors use disinformation.
  - Social media platforms a means of disseminating disinformation.
- Machine Learning/Artificial Intelligence techniques for:
  - Identifying false information.
  - Predicting the spread of information.
  - Predicting who will adopt information.
- However:
  - Complex social system with many interacting factors.
  - Adversaries are changing tactics.
  - We can’t (ethically) experiment with the real world.
  - We have limited ground truth.
  - Environment is changing.
    - Dataset shift problem.

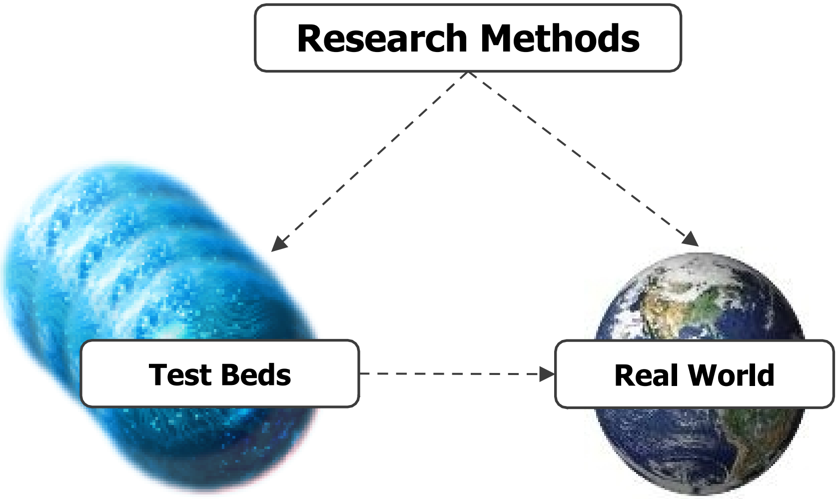


Source: FirstDraft, The essential guide to understanding the information disorder, 2019.



We are investigating the use of social simulations as a testbed.

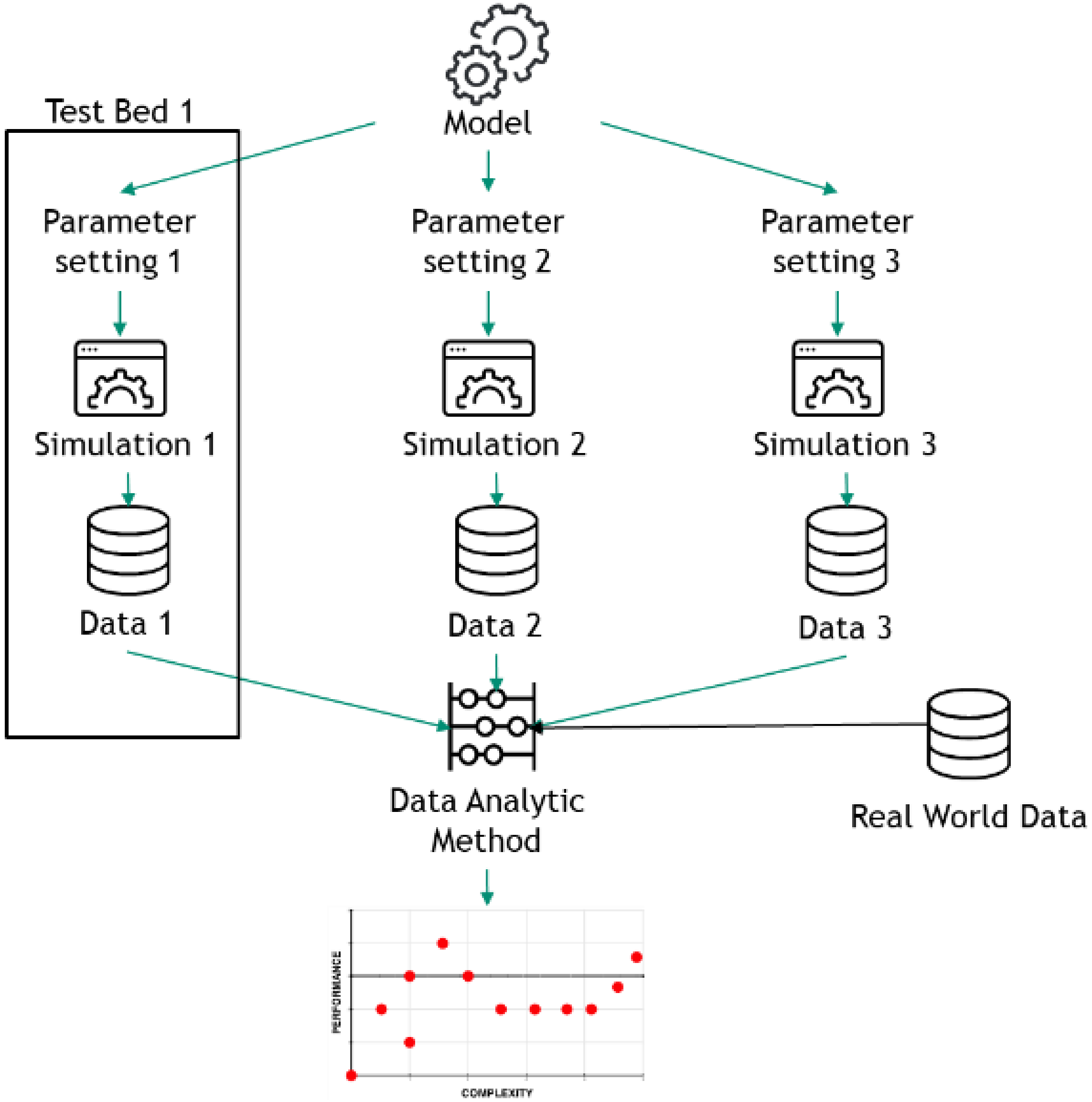
- Our approach: Use social simulations as a proxy for the real world.
- Social simulations are computational models of real-world phenomena.
  - Methods include agent-based modeling, systems dynamics, ...
- Often used for better understanding a phenomena and testing interventions in a virtual world.
- Simulations can help solve some of the problems:
  - Full ground truth.
  - Can control data bias.
  - Can run experiments and counterfactuals.
  - Can evaluate performance on varied models, parameterizations, etc.



How does the complexity of the environment impact the learnability and generalizability of ML models?

Our method requires the pairing of simulations and analytic algorithms.

Simulations vary in complexity

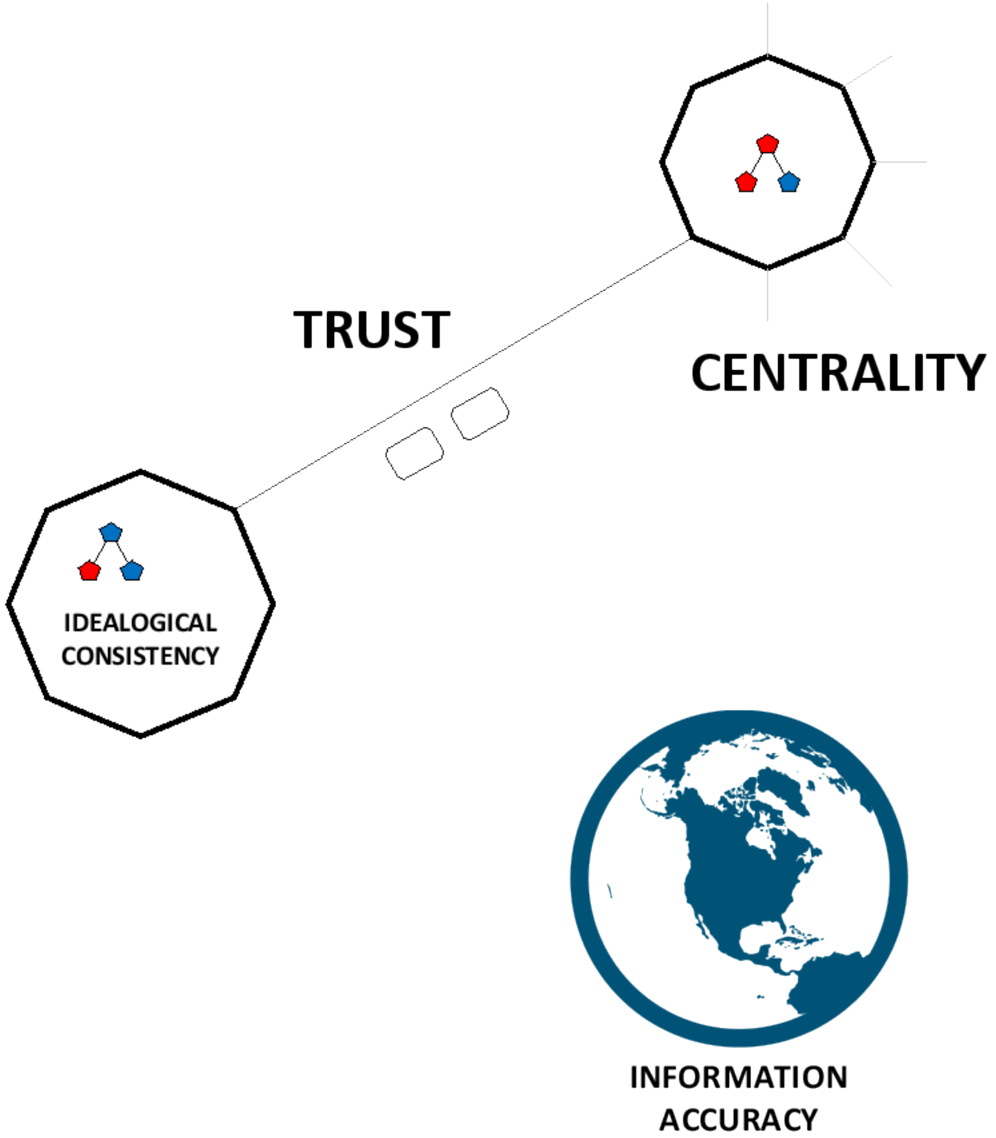


Modeling Process

- Create a simple agent-based modeling framework for person-to-person communication to **generate cascade data.**
  - Can adapt to various theoretical additions at the agent-, network-, or message-level
- Identify social-psychological theories and research with bearing on misinformation which could be added to the simple model framework
- Implement mathematical interpretation of the theories
- Identify generalizable parameter settings
  - Grid search
  - Bayesian search minimizing ABM difference to real-world cumulative distribution of retweets (Lu et al., 2014)

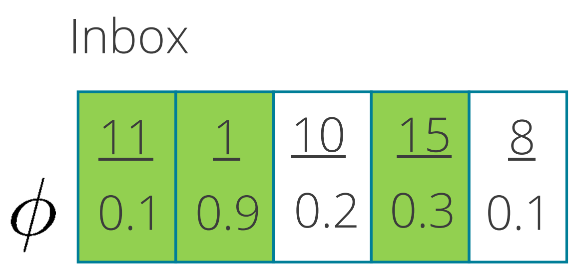
We developed a model to simulate cascades

- A cascade refers to the propagation of a piece of information (Zhou, 2021).
  - Tweet.
  - Facebook post.
  - Meme.
- Captures:
  - Send characteristics:
    - Centrality, Trust
  - Message characteristics:
    - Innate virality
    - Information accuracy
  - Receiver characteristics:
    - Trust
    - Ideological consistency



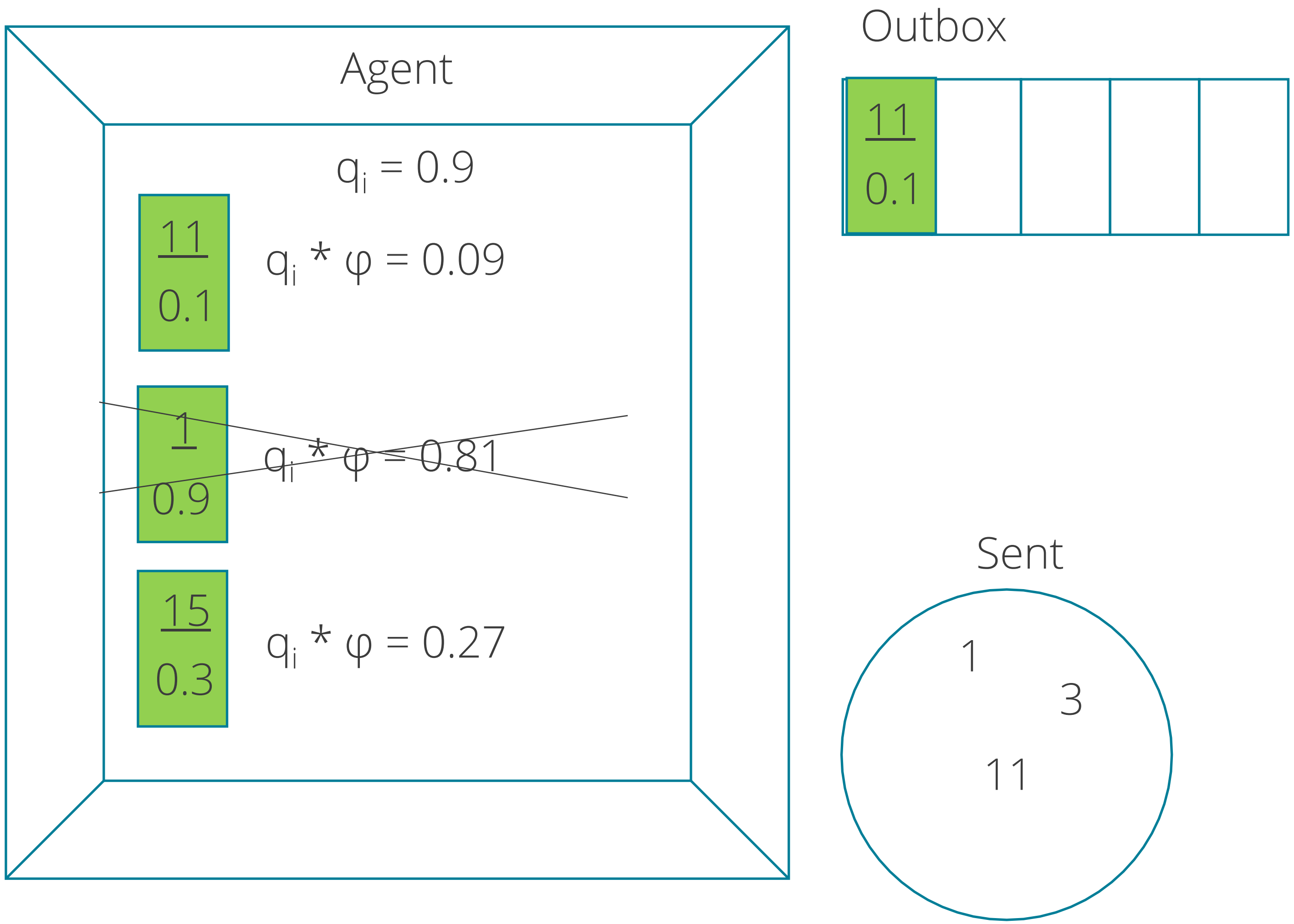
Simple Information Diffusion Model

Time: 1



K = 3

- Capture attentional constraints (k_i).
- Capture innate virality of messages (φ).
- Captures subjective likelihood to resend (q_i).



Cascade popularity prediction problem

- Given V_c^t as the number of nodes that adopted cascade c by time t.
- Goal is to predict:

$$\Delta s_c = |V_c^{t+\Delta t}| - |V_c^t|$$

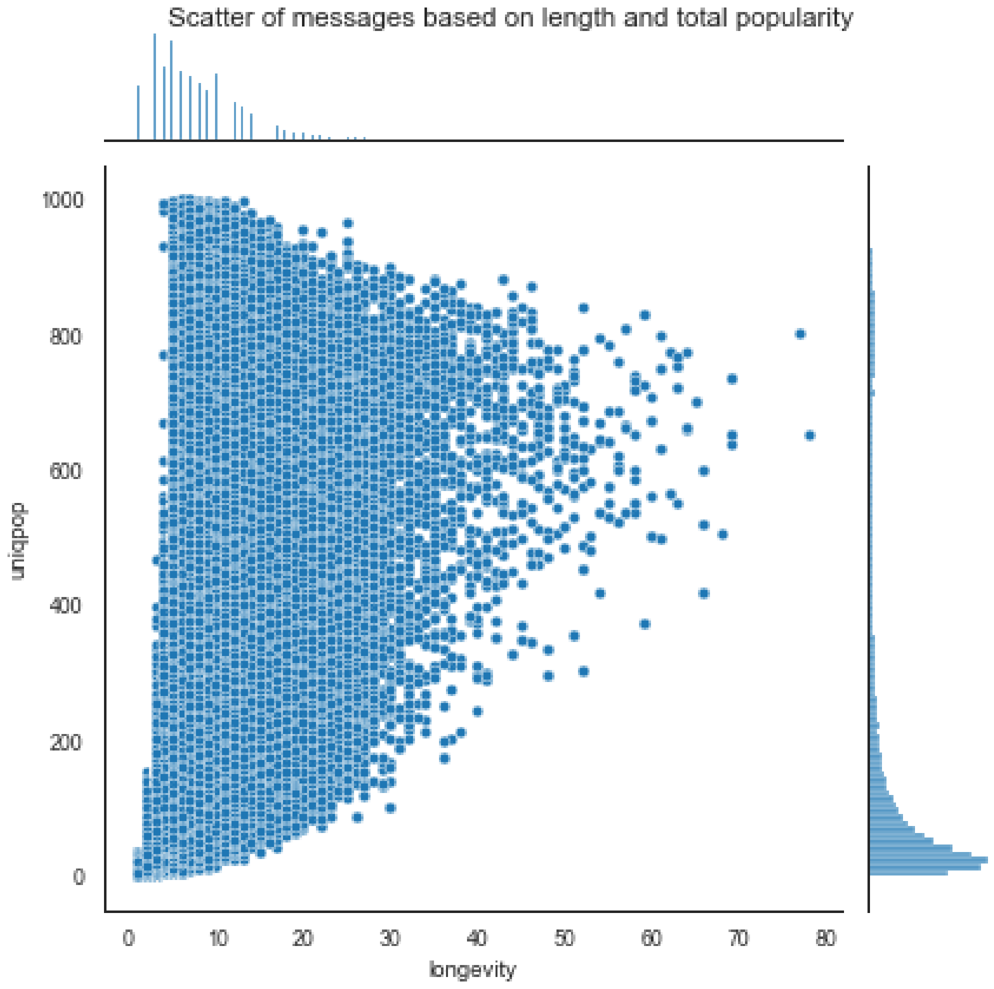
- i.e., predict the additional number of adoptions that occur within timesteps.
- Identify (dis) information that will become popular – conduct interventions.

How did we generate the training data and assess generalizability?

- Find all cascades that had at least one transmission
  - Cascade length >= 2.
- Goal: predict the change in cascade size between times 3 and 13 (i.e., t=3, \deltat=10).
- Used simple structural features of the cascades and social network.
- 80%/20%/20% partition for training/testing/validation.
- Trained each ML model on data from a simulation (a) and then tested on data from simulation (b).
- Generalizability measure: How well the ML model performs on over all simulations b.
- Three ML methods:
  - Linear Regression (Ridge)
  - Decision trees
  - DeepCas (Li, 2017)
- Error: Root Mean Squared Error (regression question).
  - abRMSE – the RMSE when an ML model trained on simulation a is applied to simulation b

We generated thousands of synthetic message cascades.

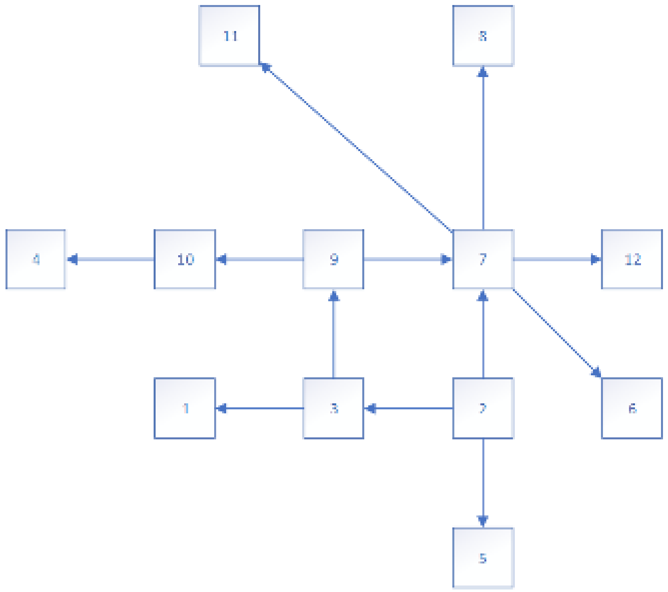
- 512 simulations with unique parameter settings.
- Each simulation has many cascades.
  - New messages were randomly seeded to agents.
- Simulations were run with 1000 agents, for 100 timesteps.
  - Intuitively think of each timestep as roughly 12 hours.
- 362,213 total cascades.
- Messages varied in length (longevity) and total number of agents reached in the social network.
- Three types of social networks:
  - Scale-free.
  - Small world.
  - Random graph.



Simple structural features of the cascades.

- Feature generation focused on simple and widely applicable structural features.
- Mean Degree and Harmonic Centrality (over all nodes in cascade, centrality measured on global graph).
- Std. Deviation of Degree and Harmonic Centrality.
- Number of components in global graph.
- Percentage of total nodes in the largest component of the global graph.
- Number of nodes in the global graph.
- Number of edges in the global graph.
- Number of nodes in the cascade.
- Number of unique nodes in the cascade (a message may be sent to the same node multiple times).

Cascade



Global Graph

In general an ML model performs worse on novel simulations.

		meanabRMSE	Std abRMSE	
Decision Tree	a != b	441	171	* Statistically significant difference at p=0
	a = b	338	174	
Linear Regression (ridge)	a != b	1,052,369	14,974,690	* Statistically significant difference at p=0
	a = b	433	1,195	
DeepCas	a !=b	5.86	13.41	* Statistically significant difference at p=0
	a = b	2.97	1.52	

Welch's t-test (allows for unequal variance and sample sizes)

Several questions..

- What were the characteristics of simulations that were the hardest to learn?
- What were the characteristics of the simulations that were the most difficult to predict?
- What were the characteristics of simulations that provided the best generality?

Which parameters caused the most difficulty in learning?

- For decision trees:
  - Er_p, sw_p, sw_k caused increases in the training RMSE.
- For linear regression:
  - Sw_k, sw_p, kimean caused increases in the training RMSE.
    - Kimean parameter that controls the attention level.
- For DeepCas:
  - sw_k, er_p, sw_p cased increases in the training RMSE.

	modelname	meanTrainRMSE	param	paramValue
25	linreg	673.605505	sw_k	40
31	linreg	602.130371	sw_p	0.5
45	linreg	499.681463	kimean	15

	modelname	meanTrainRMSE	param	paramValue
22	dtree	424.073418	sw_k	40
28	dtree	413.254332	sw_p	0.5
33	dtree	408.459125	er_p	0.008

	modelname	meanTrainRMSE	param	paramValue
25	DeepCas	3.993555	sw_k	8
43	DeepCas	3.944736	er_p	0.008
35	DeepCas	3.564137	sw_p	0.5

Which simulations were the most difficult to predict?

- For linear regression, changing the network parameters caused the most issues.
  - Sw_k, er_p, sw_p
- For decision trees, network, agent and system parameters increased difficulty.
  - Sw_k, kimean, meanAddMessages
- For DeepCas network and system parameters increased difficulty.
  - Er_p, meanAddMessages, numberofseedmessages

	modelname	meanabRMSE	param	paramValue
20	linreg	1.965870e+06	sw_k	8
32	linreg	1.594256e+06	er_p	0.008
26	linreg	1.438633e+06	sw_p	0.1

	modelname	meanabRMSE	param	paramValue
18	dtree	502.754111	sw_k	40
39	dtree	485.698618	kimean	15
6	dtree	484.097550	meanAddMessages	10

	modelname	meanabRMSE	param	paramValue
37	DeepCas	11.332321	er_p	0.008
75	DeepCas	6.663446	meanAddMessages	10
67	DeepCas	6.587822	numberofseedmessages	250

Which simulations had the best generalizability?

- Network parameters were the most influential for generalizability.

	modelname	meanabRMSE	param	paramValue
19	linreg	640.443413	sf_m	20
18	linreg	8449.026822	sf_m	4
23	linreg	51342.703364	sw_k	0

	modelname	meanabRMSE	param	paramValue
16	dtree	342.115547	sf_m	20
34	dtree	409.924353	er_p	0.04
7	dtree	413.867527	meanAddMessages	50

	modelname	meanabRMSE	param	paramValue
17	DeepCas	3.852589	sf_m	20
43	DeepCas	4.450799	er_p	0.008
24	DeepCas	4.568211	sw_k	0

Scale free and small world network topology influenced how well the model performed.

trainmodelname	sf_m		
DeepCas	0	6.391183e+00	1.522729e+01
	4	4.712620e+00	5.909841e+00
	20	3.852589e+00	2.107263e+00
dtree	0	4.598030e+02	1.639633e+02
	4	4.290900e+02	1.561458e+02
	20	3.421155e+02	1.930895e+02
linreg	0	1.397573e+06	1.726043e+07
	4	8.449027e+03	1.872571e+04
	20	6.404434e+02	1.358980e+03

trainmodelname	sw_k		
DeepCas	0	4.568211e+00	6.470345e+00
	8	8.141030e+00	2.276018e+01
	40	6.178702e+00	1.041257e+01
dtree	0	4.148034e+02	1.678334e+02
	8	5.099562e+02	1.701329e+02
	40	4.254487e+02	1.607928e+02
linreg	0	5.134270e+04	2.477342e+05
	8	1.395473e+05	3.714492e+05
	40	3.955032e+06	2.972723e+07

Conclusions

- Disinformation is a complex problem.
  - National security relevant problems have many of the same issues:
    - Complex interdependencies
    - Lack of data and ground truth.
    - Adversarial setting.
- Social simulations can serve as a testbed:
  - Full ground truth.
  - Can control data bias.
  - Can run experiments and counterfactuals.
  - Can evaluate performance on varied models, parameterizations, etc.
- Simulation parameters can change how ML models perform.
  - Training on scale-free networks provides the most generalizability.

References

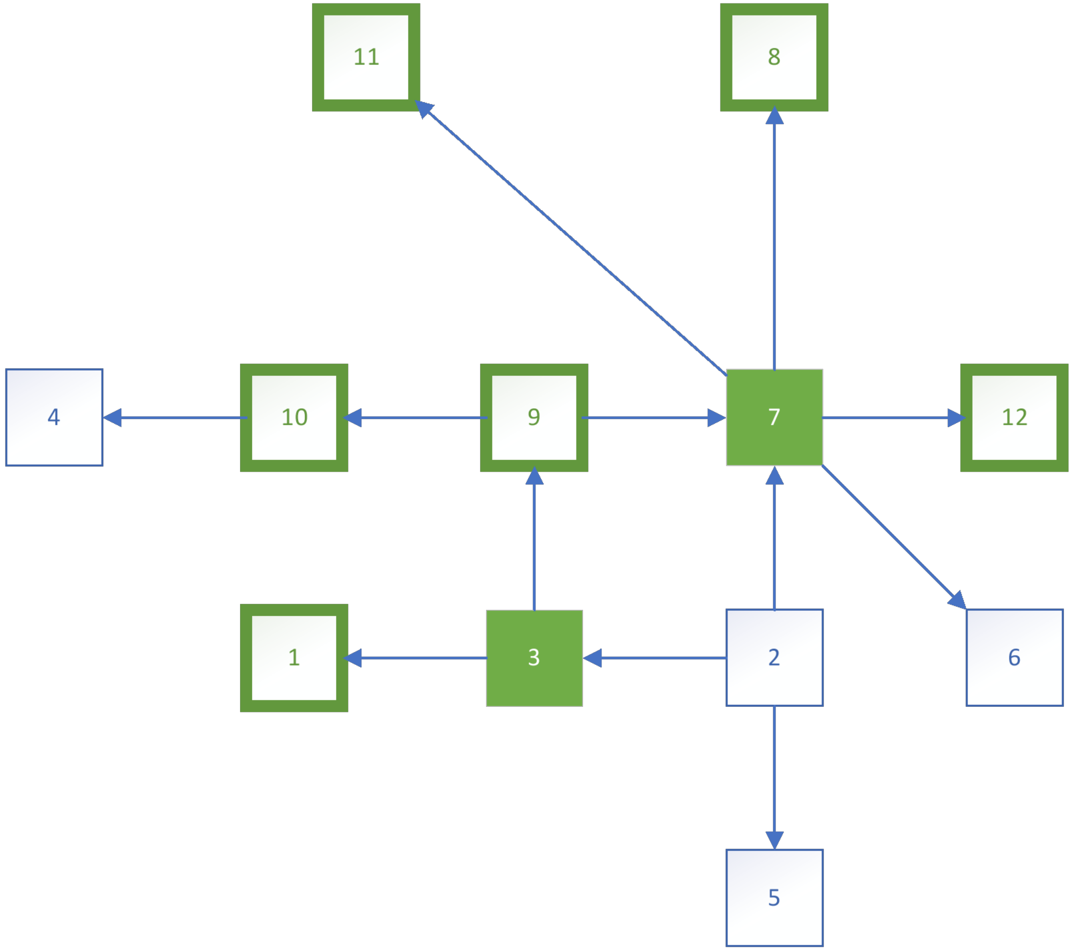
- Bradshaw S, Bailey H, Howard PN (2021) Industrialized Disinformation: 2020 Global Inventory of Organized Social Media Manipulation. Programme on Democracy & Technology, Oxford, UK
- Nguyen CV, Hassner T, Seeger M, Archambeau C (2020) LEEP: a new measure to evaluate transferability of learned representations. In: Proceedings of the 37th International Conference on Machine Learning. JMLR.org, pp 7294–7305
- Pogorelov K, Schroeder DT, Filkuková P, et al (2021) WICO Text: A Labeled Dataset of Conspiracy Theory and 5G-Corona Misinformation Tweets. In: Proceedings of the 2021 Workshop on Open Challenges in Online Social Networks. Association for Computing Machinery, New York, NY, USA, pp 21–25
- Vosoughi S, Roy D, Aral S (2018) The spread of true and false news online. Science 359:1146–1151. https://doi.org/10.1126/science.aap9559
- Zhou F, Xu X, Trajcevski G, Zhang K (2021) A Survey of Information Cascade Analysis: Models, Predictions, and Recent Advances. ACM Comput Surv 54:27:1-27:36. https://doi.org/10.1145/3433000
- Zhuang F, Qi Z, Duan K, et al (2021) A Comprehensive Survey on Transfer Learning. Proceedings of the IEEE 109:43–76. https://doi.org/10.1109/JPROC.2020.3004555

References cont..

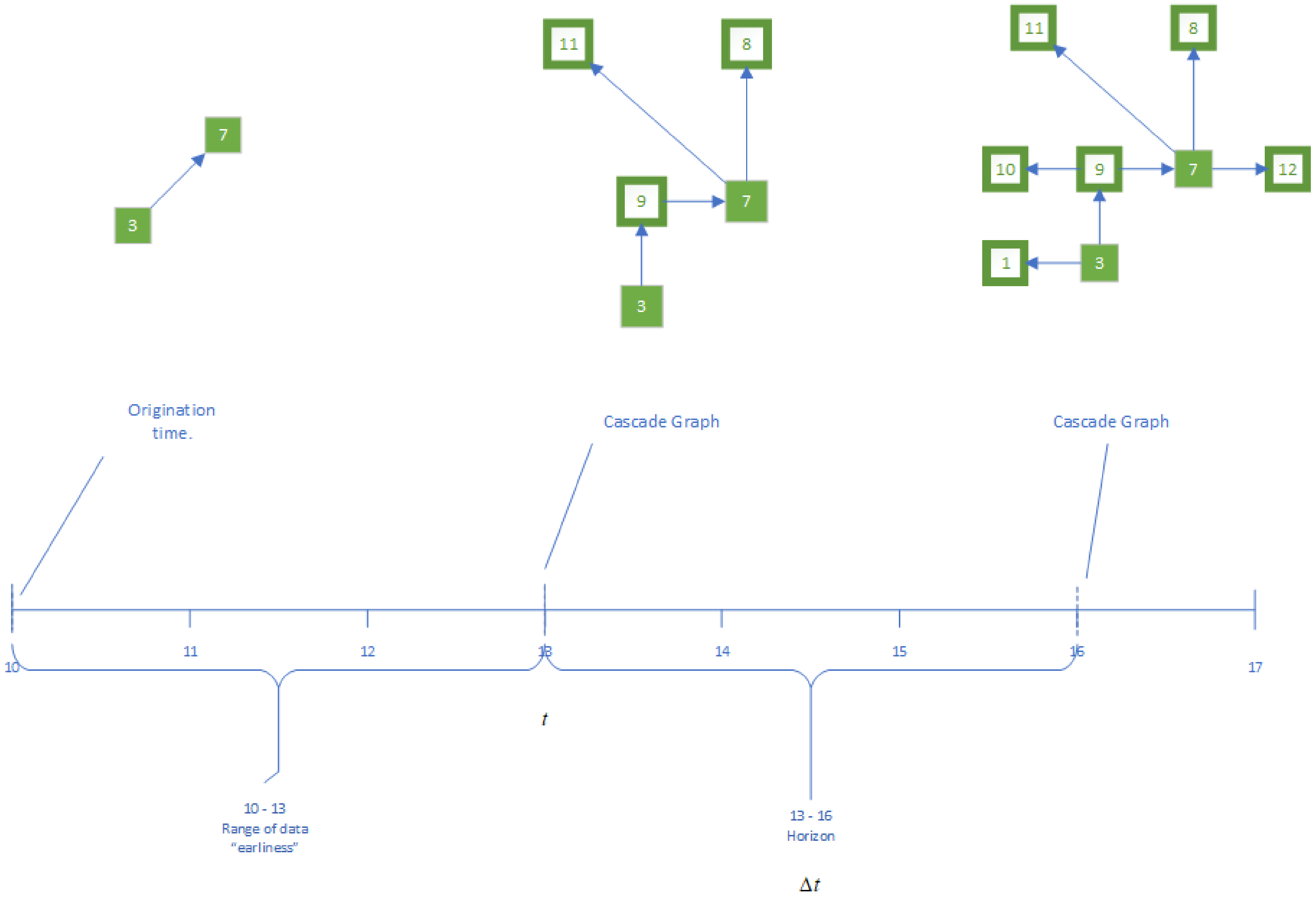
- Barbuto A, Lopolito A, Santeramo FG (2019) Improving diffusion in agriculture: an agent-based model to find the predictors for efficient early adopters. Agricultural and food economics 7(1):1–12
- Berlo DK (1960) The process of communication. New York: Holt, Rinehart, and Winston
- Fan R, Xu K, Zhao J (2018) An agent-based model for emotion contagion and competition in online social media. Physica a: statistical mechanics and its applications 495:245–259
- Festinger L (1962) Cognitive dissonance. Scientific American 207(4):93–106
- Hui C, Goldberg M, Magdon-Ismail M, Wallace WA (2010) Simulating the diffusion of information: An agent-based modeling approach. International Journal of Agent Technologies and Systems (IJATS) 2(3):31–46
- Ibarra H, Andrews SB (1993) Power, social influence, and sense making: Effects of network centrality and proximity on employee perceptions. Administrative science quarterly pp 277–303
- Igarashi T, Kashima Y, Kashima ES, Farsides T, Kim U, Strack F, Werth L, Yuki M (2008) Culture, trust, and social networks. Asian Journal of Social Psychology 11(1):88–101
- Kameda T, Ohtsubo Y, Takezawa M (1997) Centrality in sociocognitive networks and social influence: An illustration in a group decision-making context. Journal of personality and social psychology 73(2):296
- Lakkaraju K (2016) Modeling attitude diffusion and agenda setting: the mama model. Social Network Analysis and Mining 6(1):1–13
- Lewandowsky S, Pilditch TD, Madsen JK, Oreskes N, Risbey JS (2019) Influence and seepage: An evidence-resistant minority can affect public opinion and scientific belief formation. Cognition 188:124–139
- Lotito QF, Zanella D, Casari P (2021) Realistic aspects of simulation models for fake news epidemics over social networks. Future Internet 13(3):76
- Lu Y, Zhang P, Cao Y, Hu Y, Guo L (2014) On the frequency distribution of retweets. Procedia Computer Science 31:747–753
- Schweighofer S, Garcia D, Schweitzer F (2020) An agent-based model of multi-dimensional opinion dynamics and opinion alignment. Chaos: An Interdisciplinary Journal of Nonlinear Science 30(9):093,139
- Sherchan W, Nepal S, Paris C (2013) A survey of trust in social networks. ACM Computing Surveys (CSUR) 45(4):1–33
- Wang H, Zhao J, Li Y, Li C (2015) Network centrality, organizational innovation, and performance: A meta-analysis. Canadian Journal of Administrative Sciences/Revue Canadienne des Sciences de l'Administration 32(3):146–159

A cascade is the propagation of information.

- A cascade refers to the propagation of a piece of information (Zhou, 2021).
  - Tweet.
  - Facebook post.
  - Meme.
- Propagation can refer to different actions, dependent on platform:
  - Liking a tweet/post.
  - Resending a tweet/post.
- Cascade starts with initial distributors – “authors”.
- Operationally, a cascade is a sequence of users who have engaged/propagated the piece of information.
- A cascade ends after a certain time period.
- Micro-prediction: Predict the next person in the cascade.
- Macro-prediction: Predict the overall number of people in the cascade.
  - Popularity prediction.



Popularity prediction

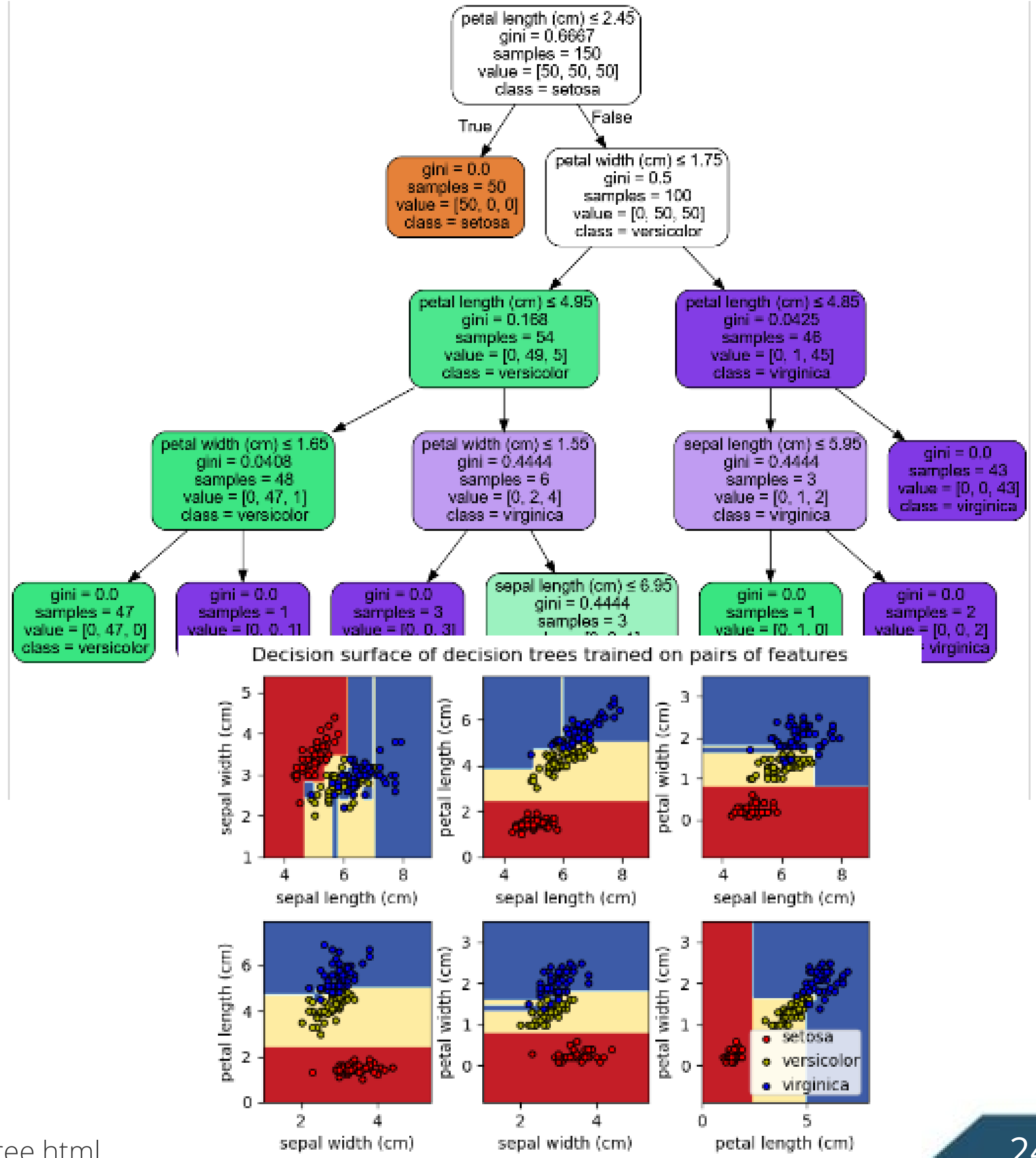


How did we generate the training data and asses generalizability?

- Find all cascades that had at least one transmission
  - Cascade length >= 2.
- Goal: predict the change in cascade size between times 3 and 13 (i.e., t=3, \deltat=10).
- Used simple structural features of the cascades and social network.
- 80%/20%/20% partition for training/testing/validation.
- Trained each ML model on data from a simulation (a) and then tested on data from simulation (b).
- Generalizability measure: How well the ML model performs on over all simulations b.
- Error: Root Mean Squared Error (regression question).
  - abRMSE – the RMSE when an ML model trained on simulation a is applied to simulation b

Three types of data analytic methods

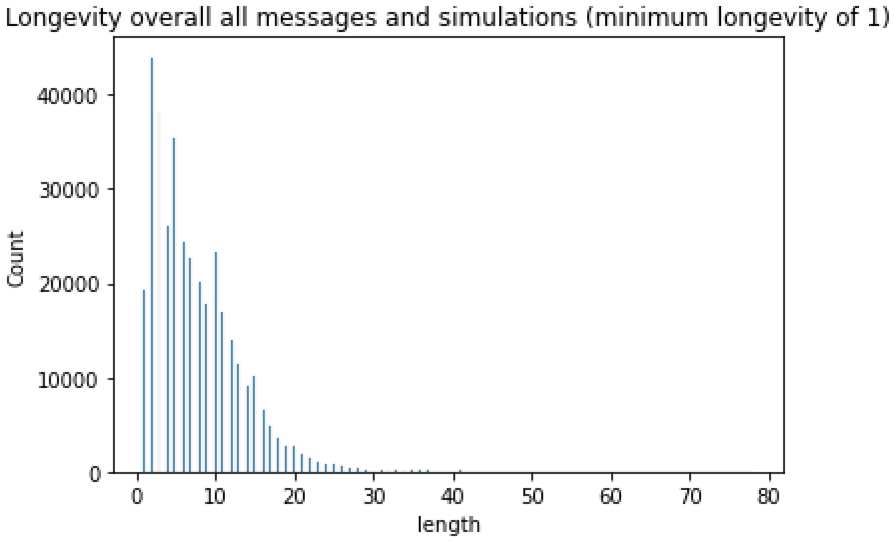
- Linear regression.
  - $y = x\beta + \varepsilon$
  - Linear combination of feature values.
  - Ridge regression (regularizes weights)
- Decision trees
  - Non-parametric method to identify decision rules on the features.
- In progress: results from a deep learning method.
- Methods were chosen to have different learning capacities.

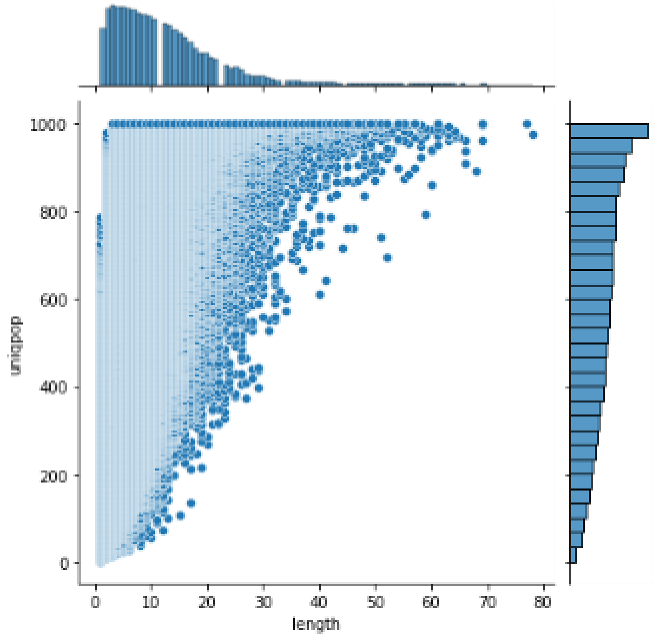


From: https://scikit-learn.org/stable/modules/tree.html

We generated thousands of synthetic message cascades.

- 512 simulations with unique parameter settings.
- Each simulation has many cascades.
  - New messages were randomly seeded to agents.
- Simulations were run with 1000 agents, for 100 timesteps.
  - Intuitively think of each timestep as roughly 12 hours.
- 362,213 total cascades.
- Messages varied in length (longevity) and total number of agents reached in the social network.
- Three types of social networks:
  - Scale-free.
  - Small world.
  - Random graph.





Next steps include comparison to real-world data.

- Our next steps include comparisons with real world data.
  - We are using the WICO (Pogorelov, 2021) dataset, which focuses on COVID misinformation spread on Twitter.
    - Example: 5G – COVID link.
- Integrating deep learning methods into the set of data analytic methods.
- Looking deeply at the impact of the parameters.

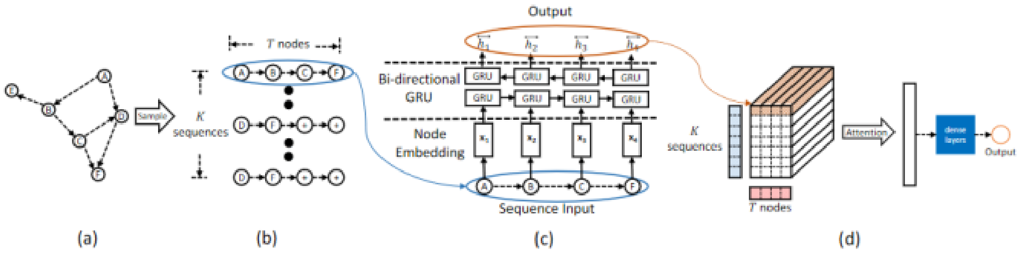


Figure 1: The end-to-end pipeline of DeepCas.

Cheng Li, Jiaqi Ma, Xiaoxiao Guo, and Qiaozhu Mei. 2017. DeepCas: An End-to-end Predictor of Information Cascades

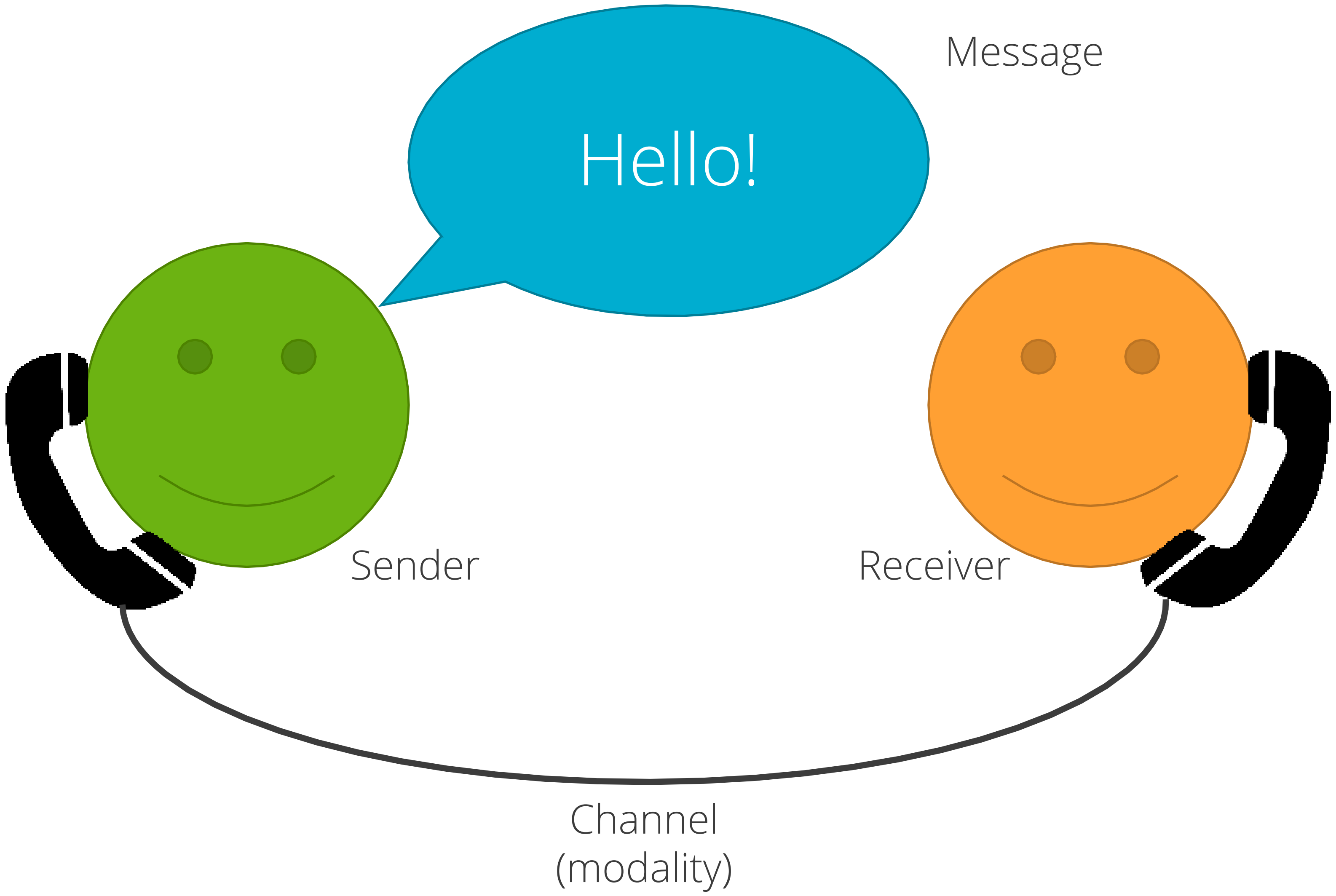
Conclusions

- Disinformation is a complex problem.
  - National security relevant problems have many of the same issues:
    - Complex interdependencies
    - Lack of data and ground truth.
    - Adversarial setting.
- Social simulations can serve as a testbed:
  - Full ground truth.
  - Can control data bias.
  - Can run experiments and counterfactuals.
  - Can evaluate performance on varied models, parameterizations, etc.

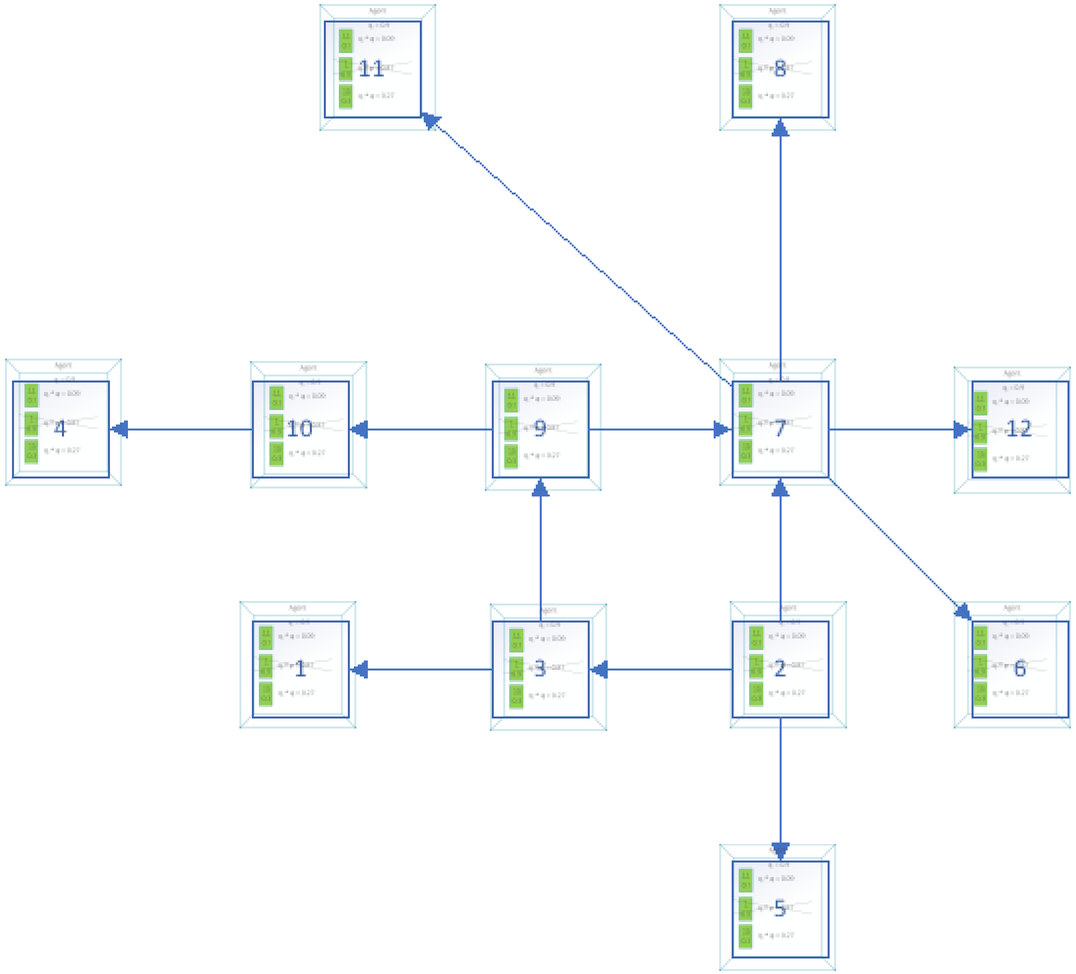
Complex Information Diffusion Model

- **Sender characteristics**
  - Credibility or authority, “speech ability” or persuasiveness, social network centrality, conformity to social norms (i.e., “Spiral of Silence”).
- **Message characteristics**
  - Topic salience, message virality, information accuracy.
- **Channel characteristics**
  - Access to communication modality.
- **Receiver characteristics**
  - Trust, cognitive/ideological consistency, “stubbornness”

Berlo (1960) SMCR model of communication



Agent model is used for each agent in a social network.



Complex Information Diffusion Model – Social Network Centrality

- **Sender characteristic** – a person's "importance" in the network, measured by their connectedness to others
  - A person's centrality is positively related with their influence on others (Ibarra et al., 1993; Kameda et al., 1997; Wang et al., 2015)
  - Centrality is operationalized in ABMs in a wide variety of ways from seeding message (Barbuto et al., 2019), to distinguishing "influencer" agents from a general public (Lotito et al., 2021)
- **In CIDM, centrality acts as a weight on inbox priority** – i.e., compared to other messages received, how likely am I to pay attention to *your* message; or how much does the algorithm weight your message compared to others
  - Eigenvector centrality, rescaled to {0:1}; model-added messages are assigned a value of 2 to ensure they are seen

Complex Information Diffusion Model – Trust

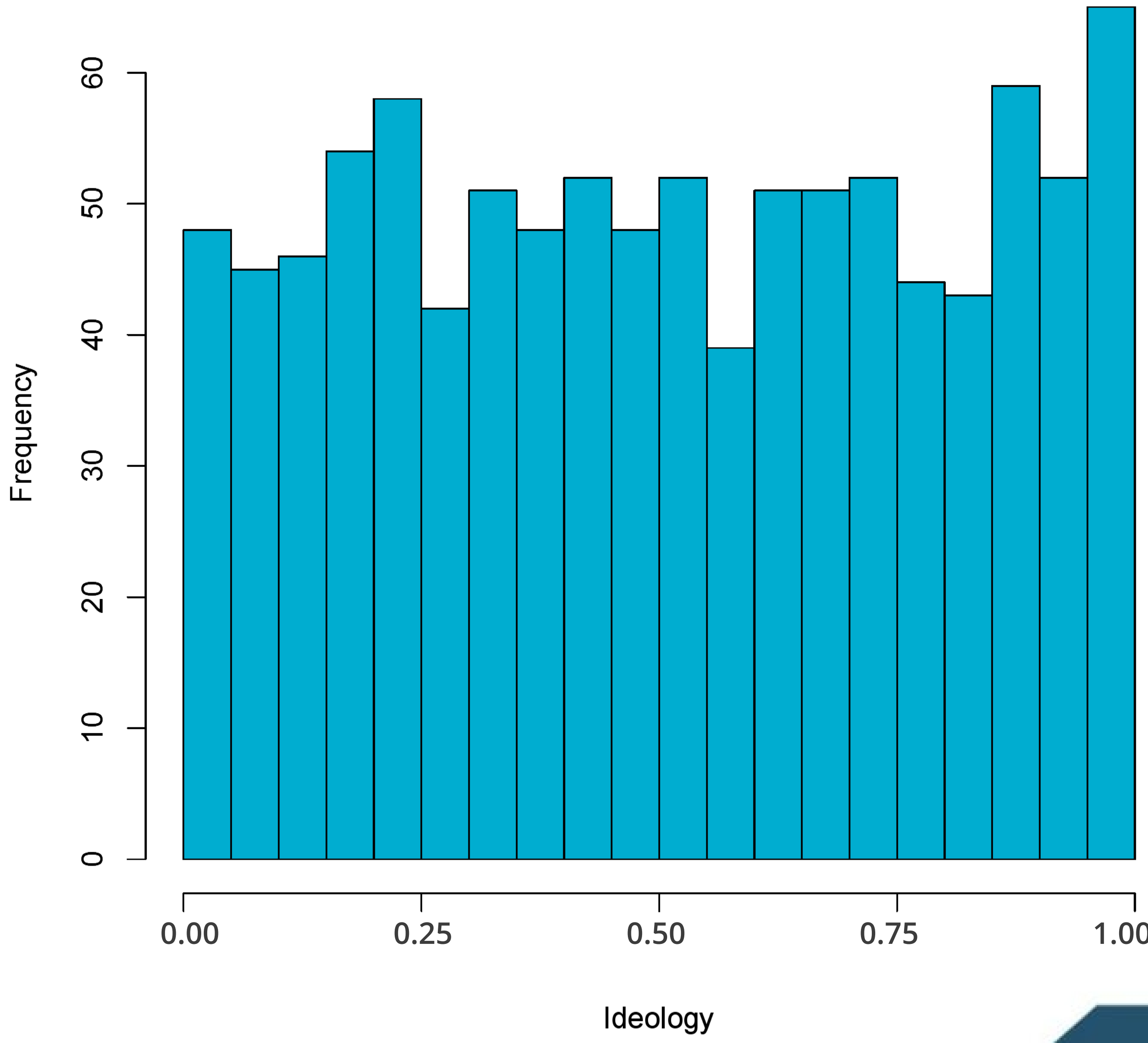
- **Directed receiver-to-sender characteristic** – a person's belief in another that the information they share is true
  - One of many aspects that affects the receiver's perception of the believability of a message, and thereby its adoption and resend probability
  - Commonly implemented as a directed edge weight in the agent-to-agent network affecting adoption and spreading rates (e.g., Hui et al., 2010); less commonly operationalized using tie reciprocity (e.g., Fan et al., 2018)
- **In CIDM, trust is an assigned directed edge value at the start of the model**; not permitted to update in this iteration
  - Can be distributed randomly, as a function of dyadic ideological similarity (Sherchan et al., 2013), or as a function of the proportion of local network overlap (i.e., triadic closure; Igarashi et al., 2008)

Complex Information Diffusion Model – Ideological Consistency

- **Receiver characteristic** – the degree to which the opinion expressed in a message on one topic aligns with the receiver's multi-dimensional ideology; greater similarity increases the probability of adopting the message, and thereby resending
  - Like cognitive dissonance theory (Festinger, 1962), but includes congruency with beliefs on other, related topics
  - Used more often in opinion dynamics models than information diffusion per se (e.g., Lakkaraju, 2016; Schweighofer, 2020)
- **In CIDM, ideological consistency increases resend probability**

Complex Information Diffusion Model – Ideological Consistency

- **Method**
  - Ideology is randomly distributed {0:1}
  - Opinions on some parameterized number of topics are drawn from a gaussian distribution with mean set at ideology, parameterized sd, and opinions beyond 0 and 1 are rounded to floor/ceiling
  - Message asserts some value in opinion space (random; {0:1}) on a particular topic
  - Consistency is 1 – mean distance of message opinion from all non-topic node opinions



Complex Information Diffusion Model

For a message (m), sent by one agent (i) to another (j), the receiving agent will resend the message with the probability:

$$P_{m \rightarrow outbox} = Virality_m * Trust_{ij} * Ideological.Consistency_{jm}$$

Complex Information Diffusion Model – Information Accuracy

- **Message/receiver characteristic** – the degree to which (receiver's perception of) information in the message conforms with (receiver's perception of) external evidence; true (or perceived true) information is more likely to be adopted and reshared
  - E.g., "vaccines are safe" message paired with evidence of few complications
  - Fairly novel in agent-based models of information diffusion, but interesting because information is modeled as both socially- and externally-supplied
  - One excellent example of its use in ABMs is Lewandowsky et al.'s (2019) model of global warming belief propagation

Complex Information Diffusion Model – Information Accuracy

- **In CIDM, information accuracy is operationalized as a filter on read messages** – perceived true information is passed through heuristic processing (trust, virality, ideological consistency), while false information is discarded
- Agents are assigned a knowledge score for each topic (variety of random distributions, {0:1})
- Each message has a random probability of being false (parameterized by topic)
- The probability of detecting that a message is false is given by a sigmoid function tied to knowledge – topic experts are more likely to accurately detect false information than non-experts

Representative Run selection

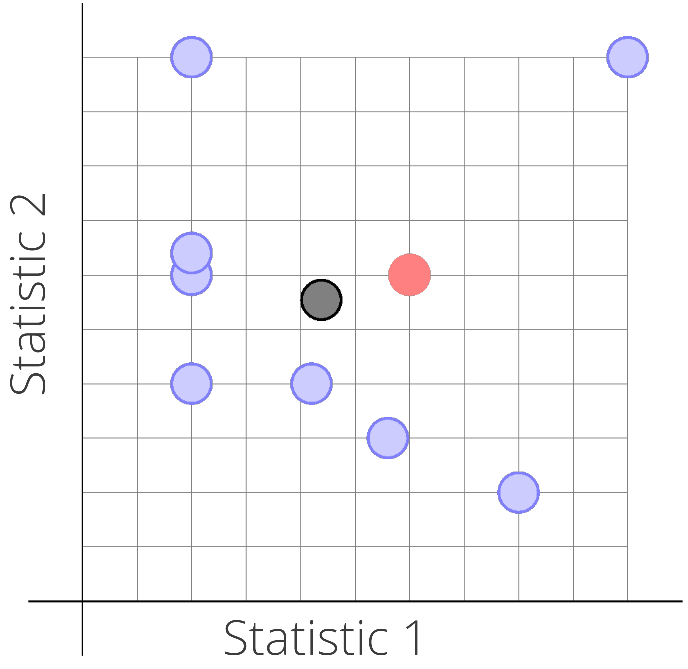


Figure 1: An example selection of the summary run. The blue dots indicate actual runs of the simulation with only the random number seed changing. The axes are values of two run characteristics. The black dot is the calculated mean along the two run characteristic dimenstions, and the red dot is the closest run to the mean.

Transferability metrics focus on the data

- Transfer learning aims to use knowledge gained in solving one task to help in another task (Zhuang, 2021).
  - Train classifier on source domain f:X -> Y.
  - How can we modify the trained f() to apply to X-> Z?
  - Example: Have a classifier for distinguishing between felines and canines. Now want to modify the classifier to learn house cats vs. leopards.
- Transferability metrics estimate how well a classifier transfer knowledge between a source domain and a target domain (Nguyen,2020).
  - Focus on the difference in the datasets.
- We are interested on the relationship between the parameters of the generative process (i.e., the parameters of the simulation) and their relation to the performance.
- Transferability metrics could be useful in assessing the similarity between simulation datasets.

Relative error

$$\omega_{i,j} = \frac{Predict(M_i, S_j^{val})}{Predict(M_i, S_i^{val})}$$

- Measures how well M_i performs on a new "world" based on trained world.
- Varies between $[0, +\infty]$.
  - If $\omega_{i,j} < 1.0$ then M_i performs better on the new data than the data it was trained on.
  - If $\omega_{i,j} \geq 1.0$ then M_i performs worse on the new data than the data it was trained on.

DeepCas: An End-to-end Predictor of Information Cascades

- DeepCas predicts size of cascade at time T in the future
- Nodes are encoded using node2vec
- Graph cascades represented as DeepWalks/Random Walks

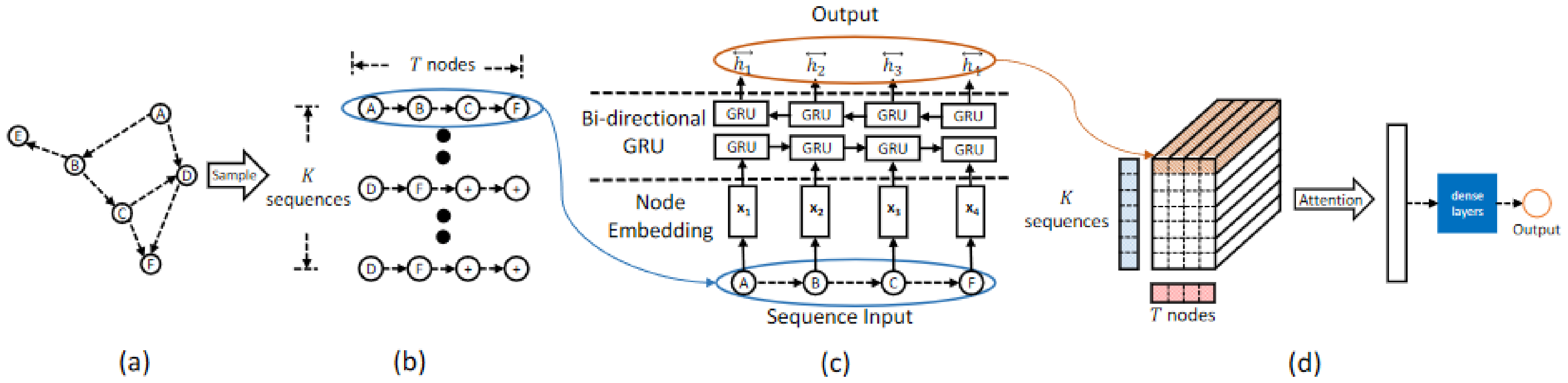


Figure 1: The end-to-end pipeline of DeepCas.

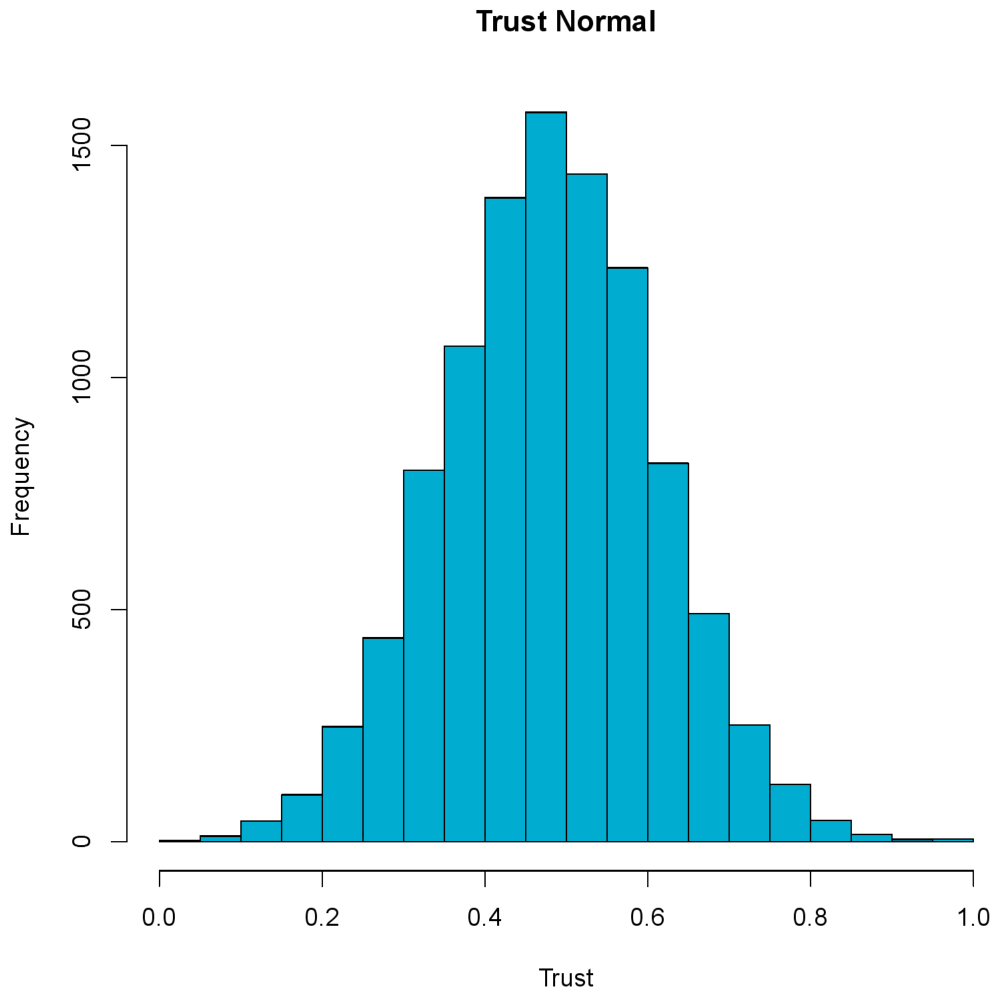
Cheng Li, Jiaqi Ma, Xiaoxiao Guo, and Qiaozhu Mei. 2017. DeepCas: An End-to-end Predictor of Information Cascades

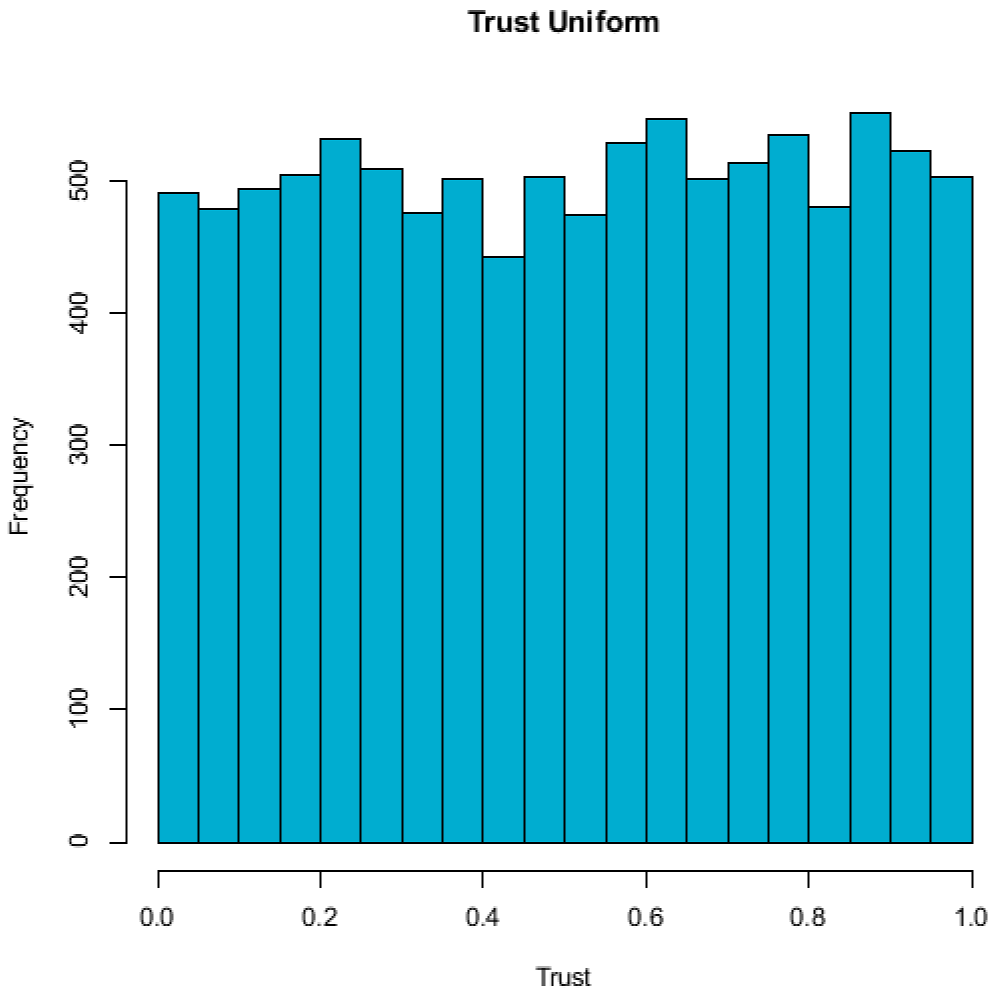
DeepCas Performance

- Average performing models were 6 units off
- Best performing model was 4.6 units off on avg

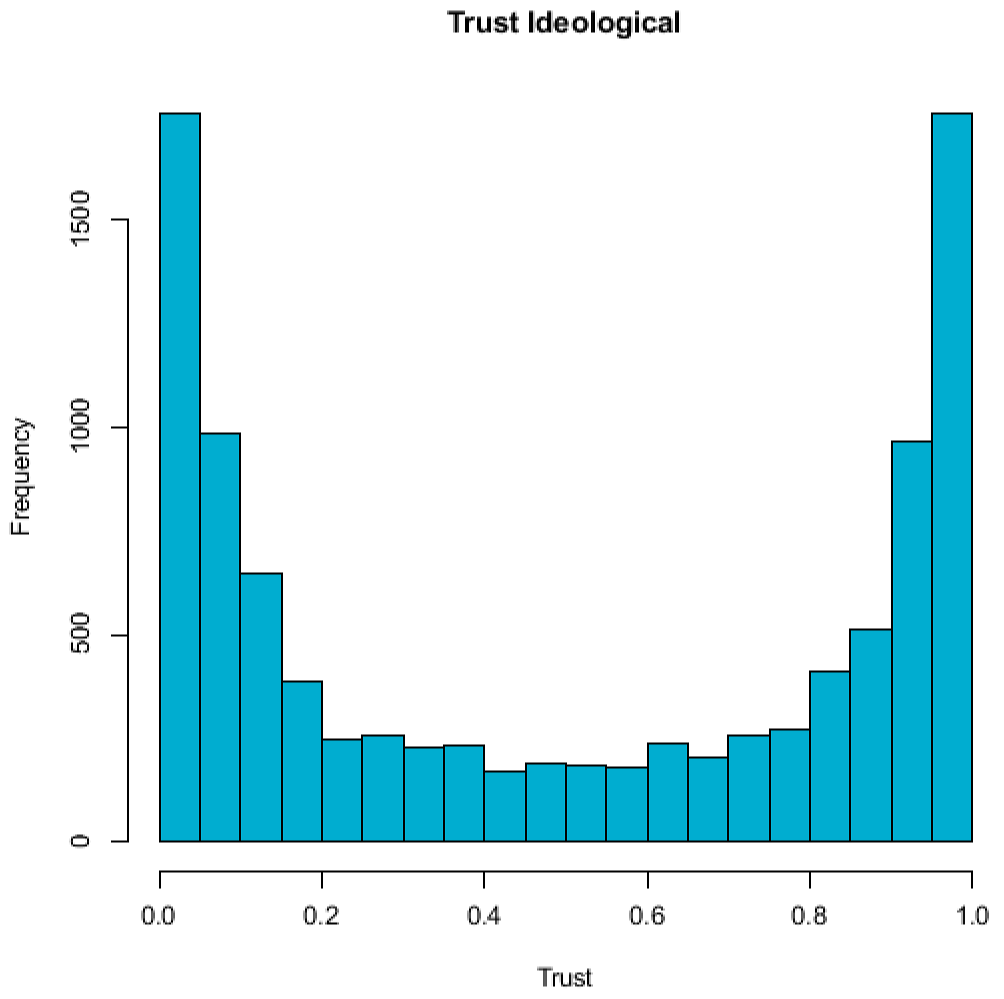
	Performance Metrics			
	Best	Median	Worst	Avg
Model Avg R2 across other datasets	-0.22453	-1.35884	-44.33005	-5.01467
Model Avg RMSE across other datasets	4.555461	5.90345	8.63677	6.61469
Model Avg Train RMSE	2.287997	4.25985	7.739711	18.79687
Model Avg Train R2	0.563479	-0.05116	-2.39109	-0.11585

Complex Information Diffusion Model – Trust

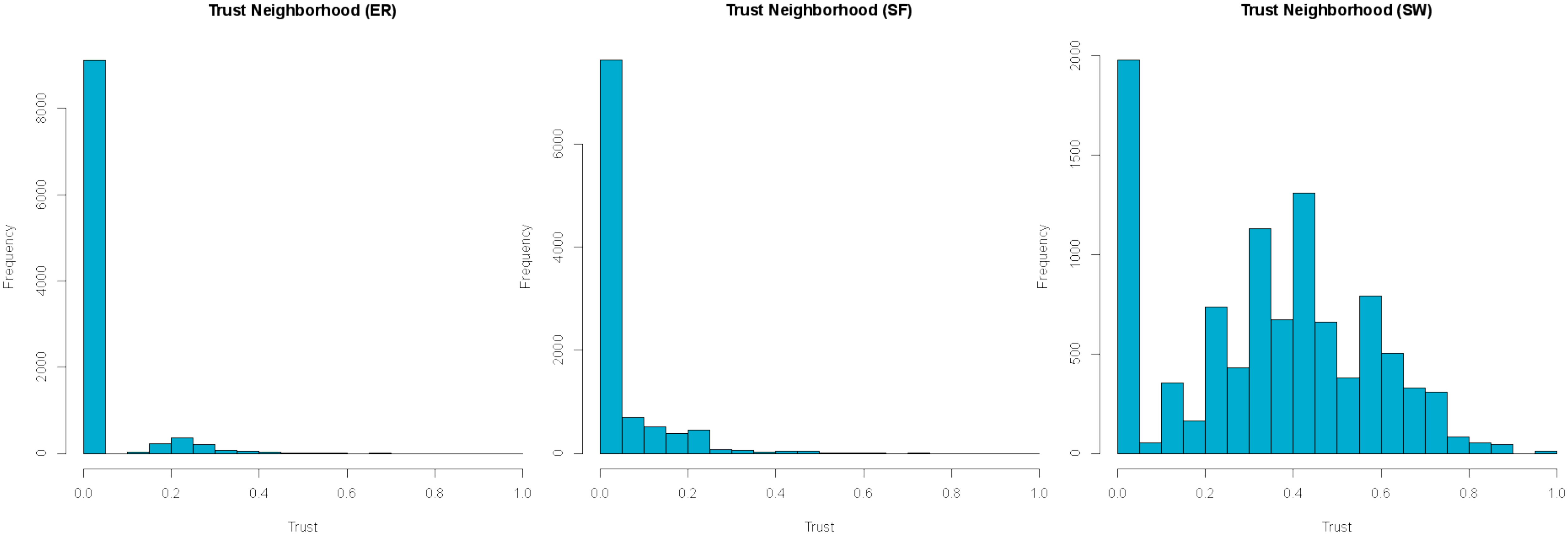




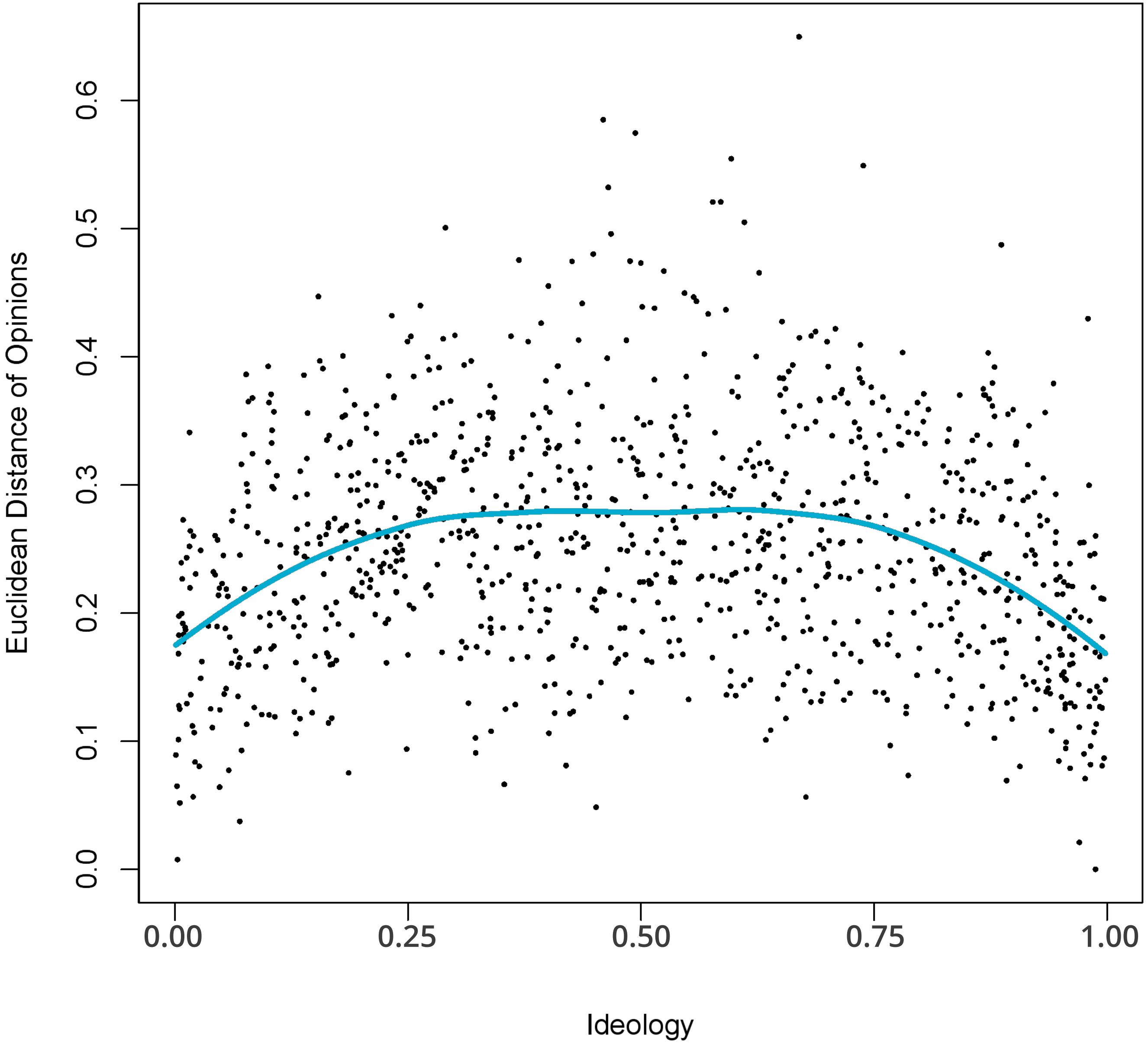
Complex Information Diffusion Model – Trust



Complex Information Diffusion Model – Trust



Complex Information Diffusion Model – Ideological Consistency



Complex Information Diffusion Model – Social Network Centrality

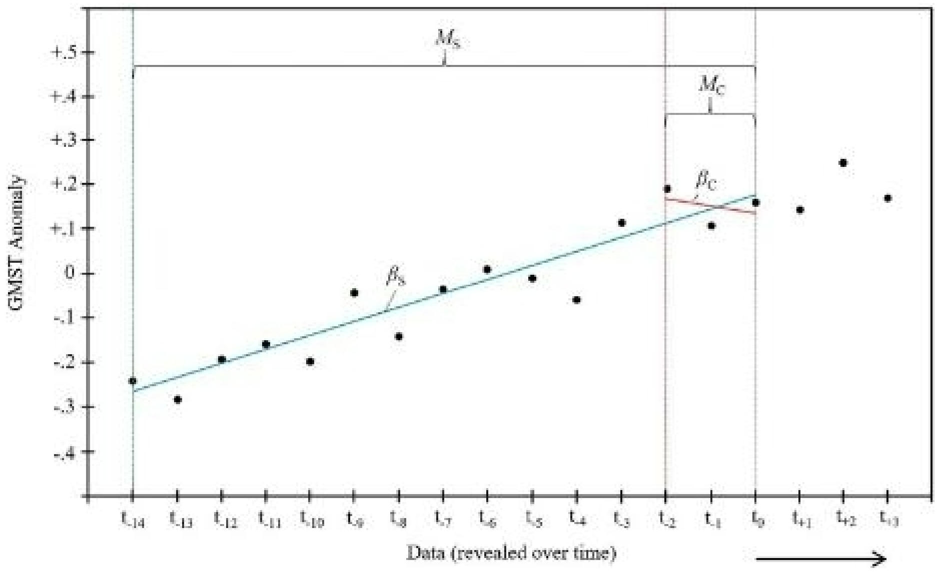
Order	Message	Sender	Centrality
1	102	i	0.83
2	103	i	0.83
3	106	j	0.55
4	102	k	0.52
5	104	l	0.11

$K_i = 3$

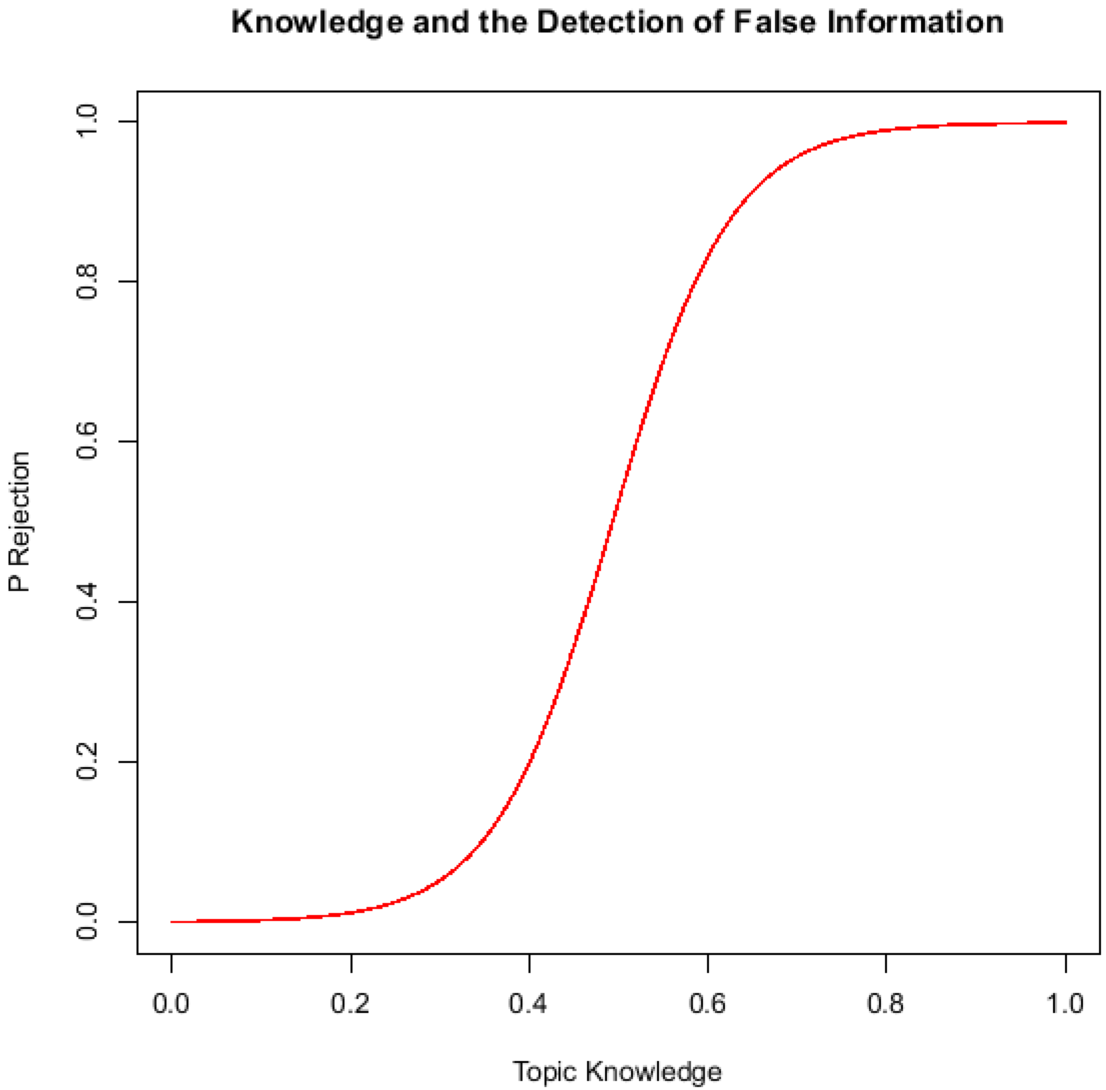
Complex Information Diffusion Model – Information Accuracy

Lewandowsky et al. (2019)

- Three types of agents: scientists, gen. pop., and contrarians
  - Varied the amount of real-world data (last 15-30 years, no data, 3 years) drawn on to form evidence-based opinion on existence of global warming; contrarians apply "skew" (see cognitive consistency)
    - Likelihood ratio drawn from linear regression slope
      - $LR = 10^{\beta - S}$
    - Bayesian belief revision
  - Scientists and contrarians confer within groups
  - They then spread to the general public 5 times per year
- Even small amounts of contrarians drastically reduce overall belief in climate change, both because of skew *and* over-reliance on small amount of data

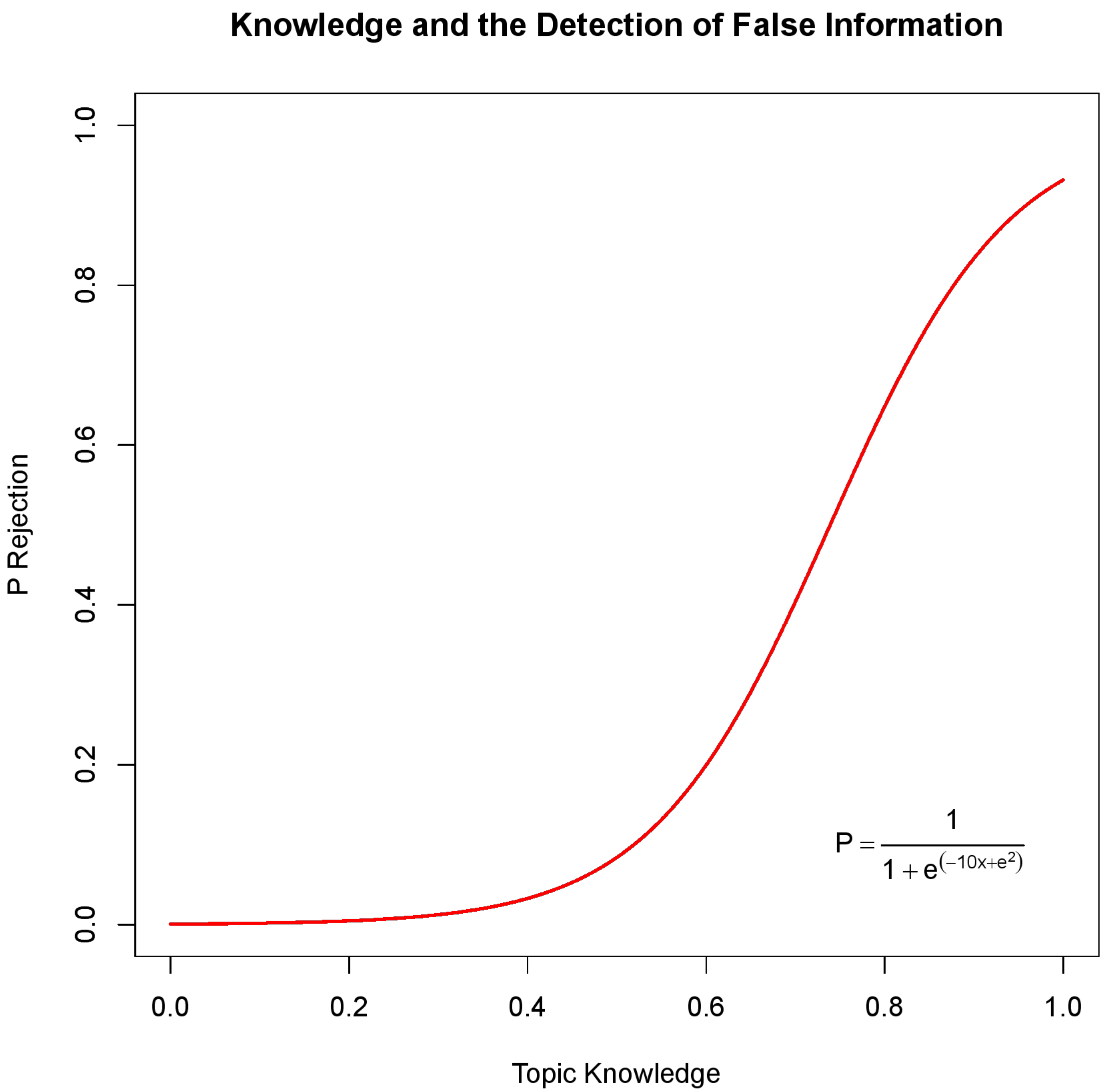


Information Accuracy

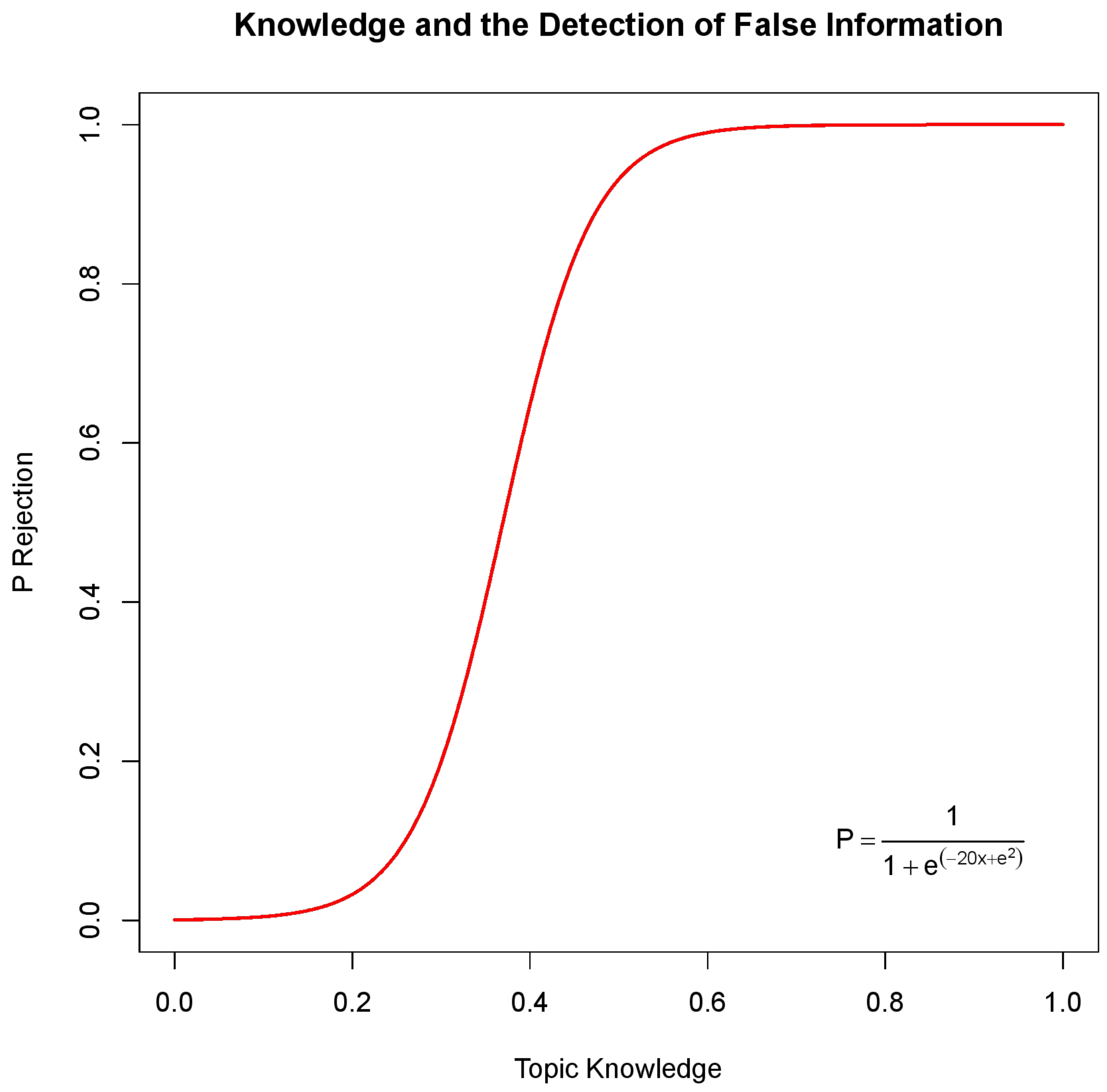


$$P = \frac{1}{1 + e^{(-15x+e^{2})}}$$

Information Accuracy



Information Accuracy



Notation to help clarify

- S_a = Data from a social simulation with parameter setting a. S_a^{train} is the training set, and S_a^{test} and S_a^{val} are the test and validation sets.
- $M_a = train(S_a^{train})$ Mi is a model trained on data generated from social simulation S_a.
- $E = Predict\left(M_a, S_b^{val}\right)$: The performance of using M_a to predict based on data generated from social simulation S_b.
  - i.e., how well does a model perform when its used on data from a different "world".
  - Note that the dependent variable may be the same, but fundamental elements of the simulation may differ between S_a and S_b.
  - For instance, one could be a situation where the social network is scale free, and Sj a situation where the social network is small world.
  - Error will be defined per problem domain/model type.

Error is calculated as RMSE

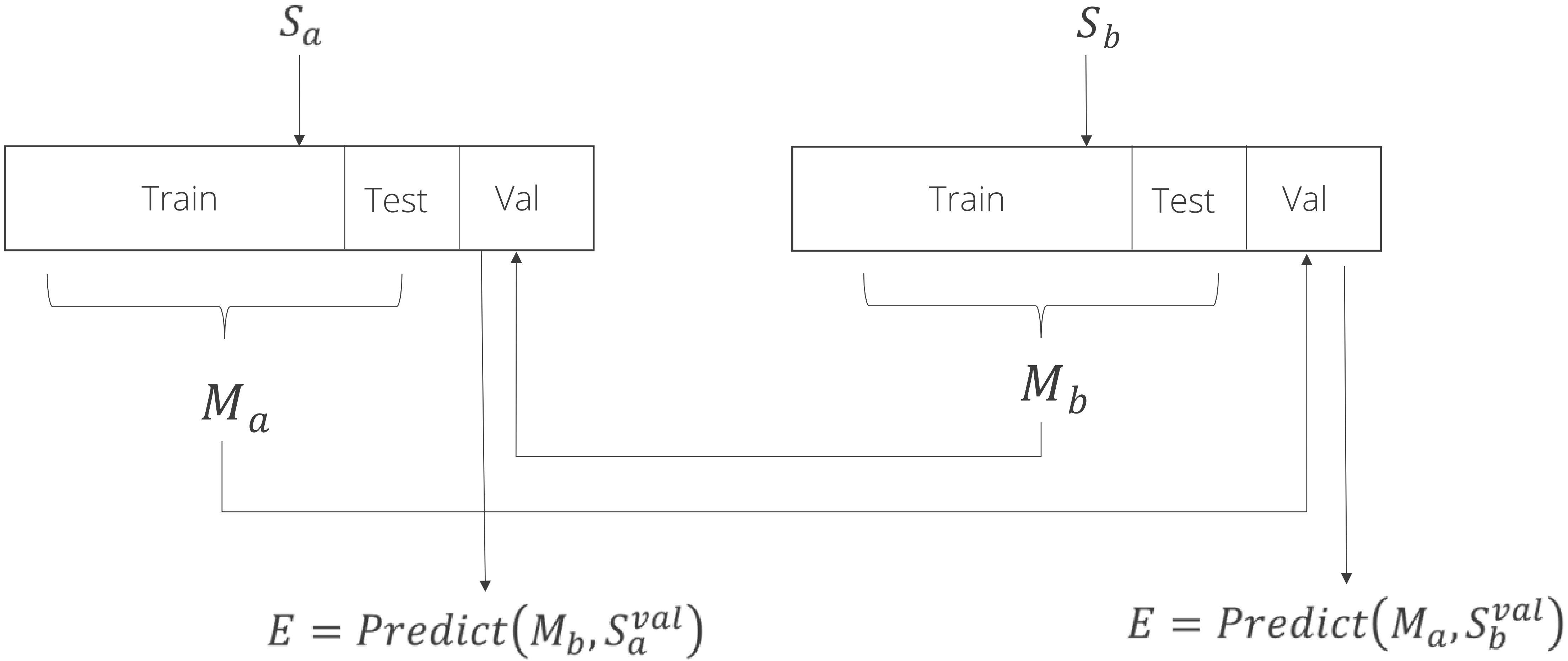
$$trainRMSE = Predict(M_a, S_a^{train}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

$$abRMSE = Predict(M_a, S_b^{val}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

$$meanabRMSE_a = \frac{1}{m}\sum_{b=1}^{m} Predict(M_a, S_b^{val})$$

Where m is the number of simulations, and n is the number of samples in S_a^{train} or S_b^{val}

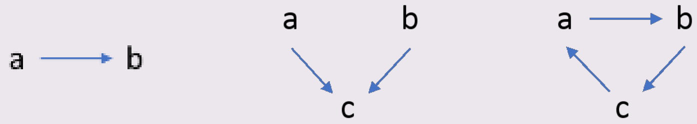
- $meanabRMSE_a$ serves as a measure of "generalizability" of a over the set of simulations b. The lower the it is, the better M_a performs on a wide variety of simulations.

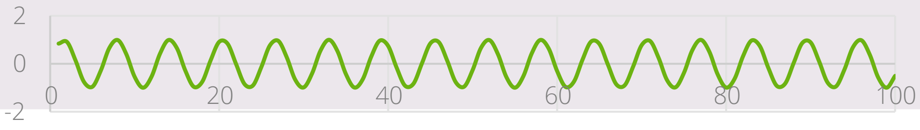


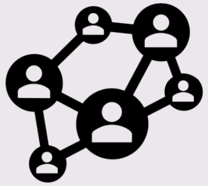
S_a
Train
Test
Val
M_a
S_b
Train
Test
Val
M_b
$E = Predict(M_b, S_a^{val})$
$E = Predict(M_a, S_b^{val})$

We focus on graph structure as a measure of complexity.

	Not tied to social sciences	Inspired by the social sciences
Requires knowledge of system structure	Measures of System Intricacy *How complicated is the causal structure?* Metric: Causal Complexity	Behavioral Capacity *How do interactions and behaviors of actors affect complexity?* Metric: Number of Differentiated Relationships
Does not require knowledge of system structure	Information-Theoretic Complexity *What is the information content of the system's behavior?* Metric: Time-Averaged Normalized Compression Distance	Measures of Social Organization *How organized are social relationships in the system?* Metric: Global Reaching Centrality





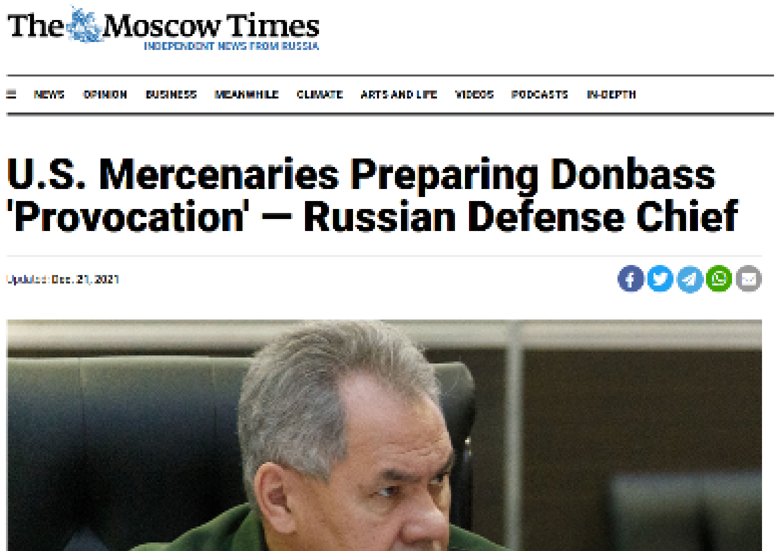


Russia launched disinformation campaigns during the 2014 Crimea conflict.

- Propaganda campaign over TV stations and social media sites.
- Spread of disinformation painting the west as fascists.
- Disinformation enhanced the divisions that were already there.

"The fundamental purpose of Russian disinformation is to undermine the official version of events — even the very idea that there is a true version of events

The disinformation launched during this campaign was not aimed to rally audiences to Russia's point of view, but to **exacerbate social tensions and plant doubt** about the presence of any empirical truth."



https://www.theguardian.com/world/2014/mar/17/crimea-crisis-russia-propaganda-media
https://crestresearch.ac.uk/resources/russia-and-disinformation-the-case-of-ukraine-full-report/

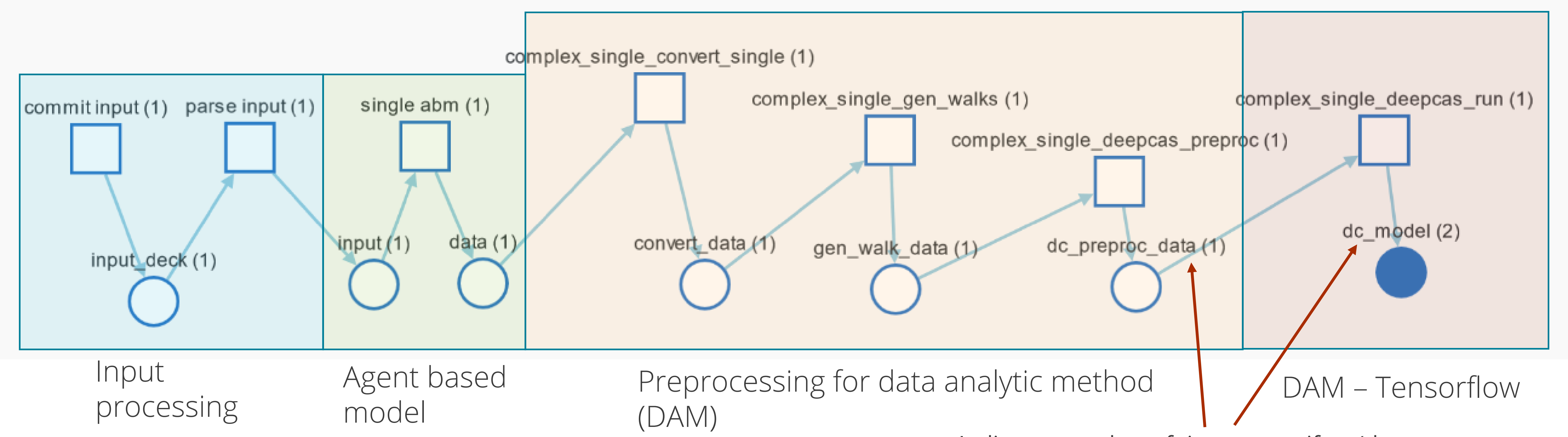
Project goal: What makes an effective simulation testbed?

- Generates data appropriate to the data science methods being evaluated.
- Can manipulate the simulation to evaluate varied scenarios, counterfactuals, etc.
- Should have clear causal relationships.
- Causal information, equations, can serve as ground truth.
- Simulation testbeds should generate non-trivial data.
- Stochasticity should be seeded.
- Should generate behavior without external manipulation.

Operationalizing the workflow using WandB

- We explored the Sandia instance of WandB and realized it can be applied as a tool for tracking, versioning and parameter sweeps beyond just ML applications.

  - Example dag from WandB (single-run version of our workflow)



- Indicates number of times an artifact (data or saved model) has been generated.
- In our case, we have thousands of data sets generated from ABM and DAM. WandB allows us to track these to their inputs.

Working code resides in our project GitHub:
https://cee-gitlab.sandia.gov/simtestbed/pipeline
WandB project space:
https://wandb-prod.sandia.gov/simtestbed

Simple Information Diffusion Model

- Agent’s have an inbox and an outbox for messages (schedule-agnostic) with some messages seeded randomly at tick 0, and others added in parameterizable intervals
- Message characteristic: virality – random {0:1}
- Agent characteristics: limit on inbox read – random integer, and individual variance in send rate – random {0:1}
- Extend this model using social and psychological processes (complex information diffusion model)

Grid Sweep Parameter Settings

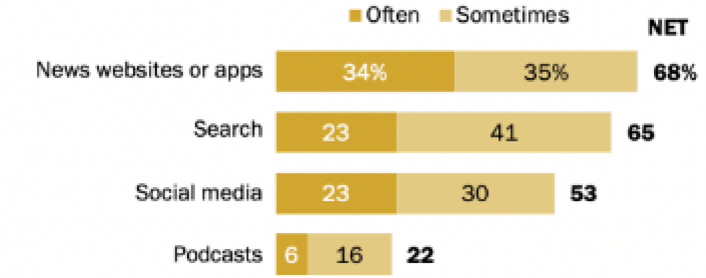
- **Number of seeded messages:** 50, 250
- **Number of agents seeded with each new message:** 50
- **Message virality drawn from power distribution with alpha:** 4
- **Number of agents:** 1,000
- **Max number of timesteps:** 100
- **Number of topics:** 3
- **Probability of false message by topic:** (0.1, 0.1, 0.1)
- **Number applied to the false detection sigmoid function by topic:** (4, 4, 4)
- **Add new messages every x ticks:** 5
- **Every x ticks, add mean(SD) messages:** 10(2), 50(10)
- **Network type:** random, scale free, small world
  - **Network density:** 0, 0.008, 0.04
  - **Small world re-wiring probability:** 0, 0.1, 0.5
- **How do distribute trust along all directed edges:** random uniform, 1-mean distance of opinions (ideological homophily)
- **Qi mean(SD) – subjective resend probability:** 1(0.2)
- **Ki mean(SD) – subjective attention limit on inbox:** 5(1), 15(3)
- **How to distribute ideology:** random uniform, random Gaussian (M = 0.3, SD = 0.2)
- **How to distribute topic opinions from ideology:** small random Gaussian (M = ideology, SD = 0.05), large random Gaussian (M = ideology, SD = 0.25)
- **How to distribute topic knowledge:** triangular distribution with mode (0.2, 0.2, 0.2)

*Highlighted parameters were varied in the grid sweep of every unique parameter combination

Americans are moving toward online and social media sources for news.

Americans more likely to get news on digital devices from news websites, apps and search engines than from social media

% of U.S. adults who get news ____ from ...

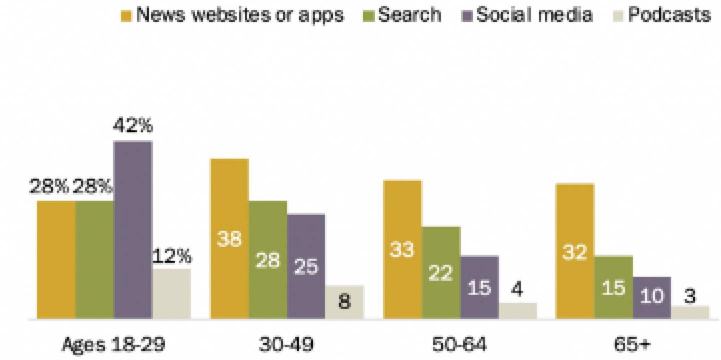


Source: Survey of U.S. adults conducted Aug. 31-Sept. 7, 2020.

PEW RESEARCH CENTER

Online, most turn to news websites except for the youngest, who are more likely to use social media

*% of U.S. adults who get news **often** from ...*

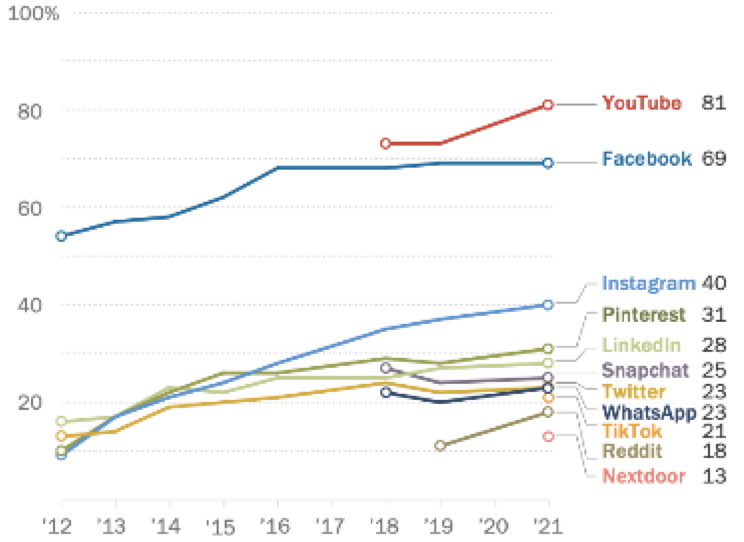


Source: Survey of U.S. adults conducted Aug. 31-Sept. 7, 2020.

PEW RESEARCH CENTER

Growing share of Americans say they use YouTube; Facebook remains one of the most widely used online platforms among U.S. adults

% of U.S. adults who say they ever use ...



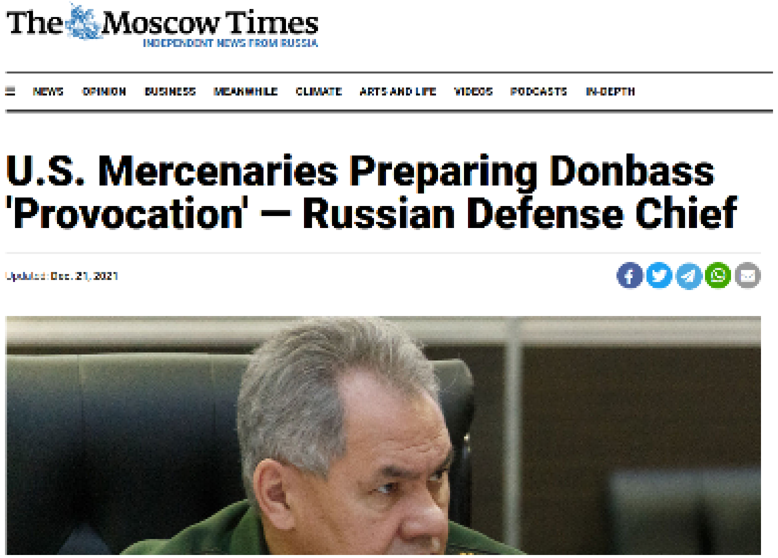
Note: Respondents who did not give an answer are not shown. Pre-2018 telephone poll data is not available for YouTube, Snapchat and WhatsApp; pre-2019 telephone poll data is not available for Reddit. Pre-2021 telephone poll data is not available for TikTok. Trend data is not available for Nextdoor.

Source: Survey of U.S. adults conducted Jan. 25-Feb. 8, 2021.

"Social Media Use in 2021"

PEW RESEARCH CENTER

https://pewrsr.ch/2MZqns7



https://www.globaltimes.cn/page/202109/1233773.shtml

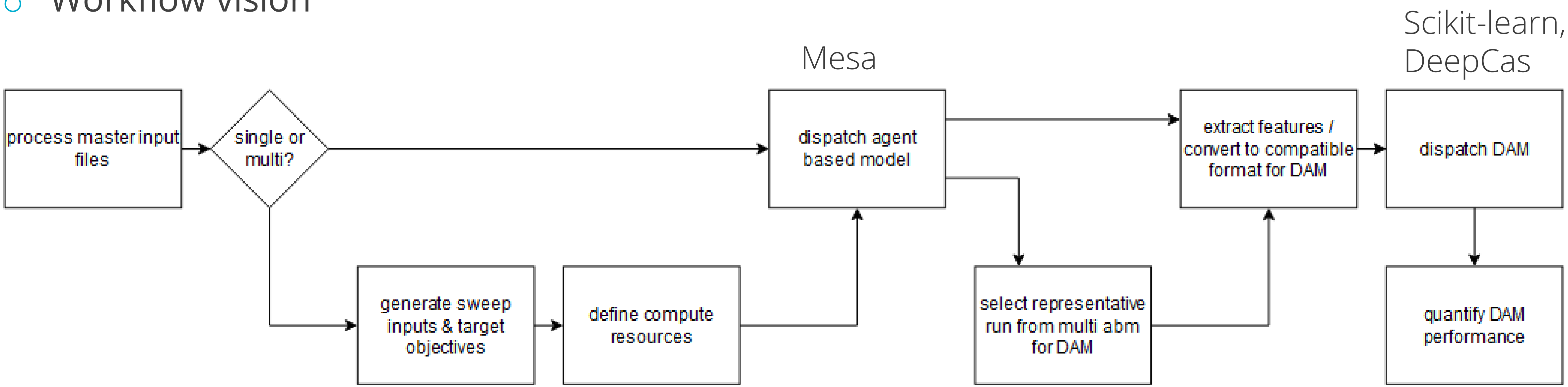
"Falsehood diffused **significantly farther, faster, deeper, and more broadly than the truth** in all categories of information, and the effects were more pronounced for false political news than for false news about terrorism, natural disasters, science, urban legends, or financial information." (Vosoughi, 2018)

A flexible, scalable infrastructure is needed.

- Challenge
  - Our problem calls for the creation of datasets of information cascades and subsequent analysis to develop a generalized description of the data that could explain real world phenomena.
  - To the best of our knowledge, no platform currently exists that combines:
    - Simulation of network information cascades to generate large volumes of synthetic data
    - Development of ML models of various complexities on both synthetic and real-world data
    - Execution within a feasible timeframe
- Solution
  - We need to develop an architecture that integrates:
    - Agent-based modeling for synthetic data creation
    - Extraction of features from synthetic and real-world data
    - Analysis of data and model development using machine-learning approaches
    - Tracking results across multiple parameter settings
    - Fast, distributed computing

Architecture of the solution.

- Workflow vision



- Executed on the Common Engineering Environment and Synapse.
- Coded in Python.
- We explored the Sandia instance of WandB and realized it can be applied as a tool for tracking, versioning and parameter sweeps beyond just ML applications.

Working code resides in our project GitHub:
https://cee-gitlab.sandia.gov/simtestbed/pipeline
WandB project space:
https://wandb-prod.sandia.gov/simtestbed

Small changes in the world can cause unknown effects

- Social media companies change user elements all the time.
  - Twitter changed the length of tweets.
  - Facebook included “frowny face” emojis.
- Social media companies change invisible elements as well.
  - Facebook and Twitter friend/follower recommendation algorithms.
  - News/message recommendation.
  - Search results (based partly on auction).
- The world changes
  - Demographics of usage shift.
  - New technologies emerge (Snapchat/Tiktok vs. Facebook).


RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES

Experimental evidence of massive-scale emotional contagion through social networks

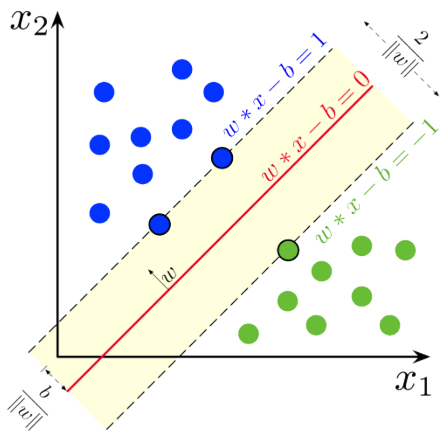
Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock Authors Info & Affiliations

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved March 25, 2014 (received for review October 23, 2013)

June 2, 2014 | 111 (24) 8788-8790 | https://doi.org/10.1073/pnas.1320040111

Data based methods can help – but have limitations..

- Machine Learning/Artificial Intelligence techniques for:
  - Identifying false information.
  - Predicting the spread of information.
  - Predicting who will adopt information.
- However:
  - Complex social system with many interacting factors.
  - Adversaries are changing tactics.
  - We can't (ethically) experiment with the real world.
  - We have limited ground truth.
  - Environment is changing.
    - Dataset shift problem.

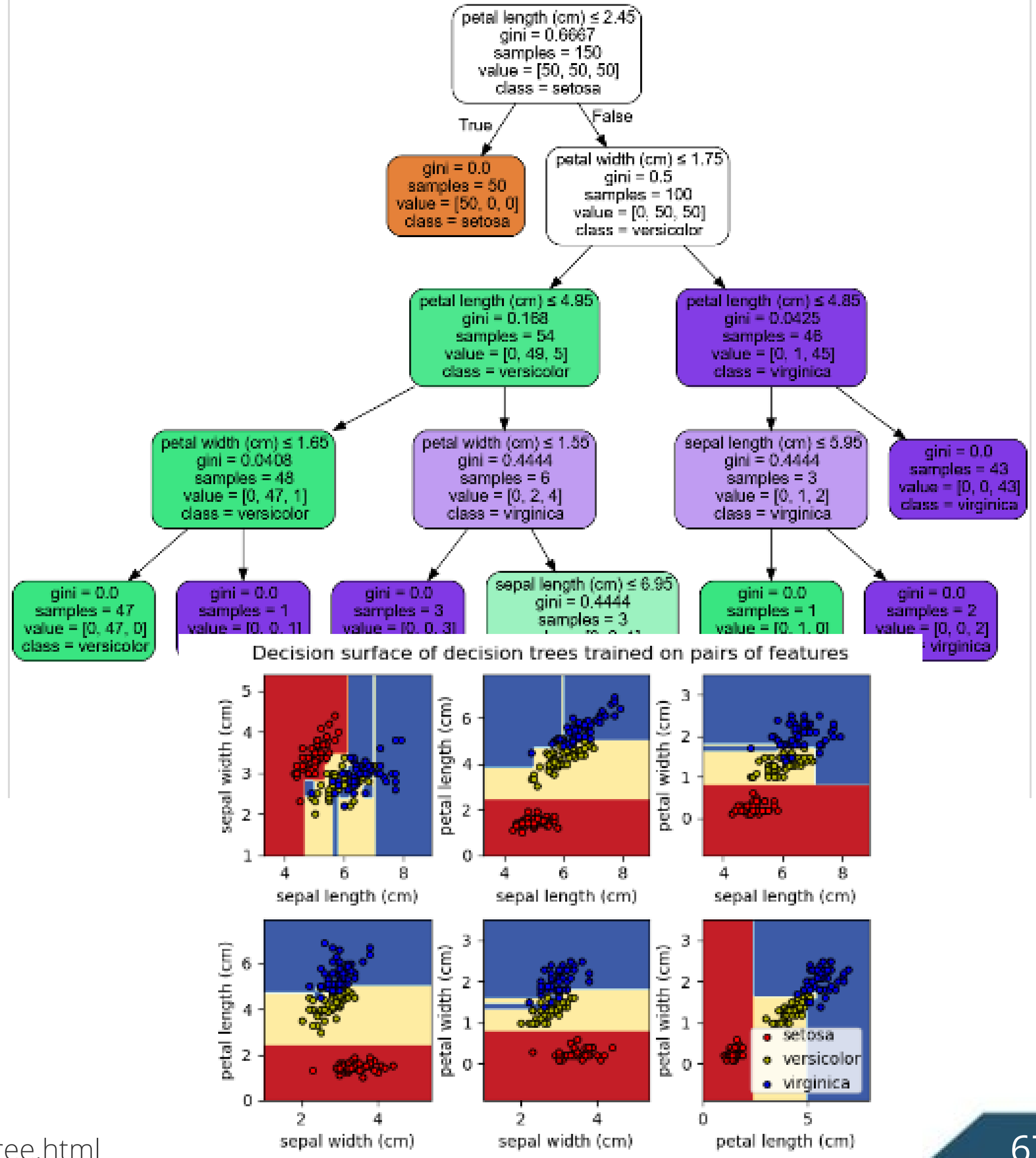


https://en.wikipedia.org/wiki/Support-vector_machine

https://www.soz.psy.unibe.ch/

Three types of data analytic methods

- Linear regression.
  - $y = x\beta + \varepsilon$
  - Linear combination of feature values.
  - Ridge regression (regularizes weights)
- Decision trees
  - Non-parametric method to identify decision rules on the features.
- In progress: results from a deep learning method.
- Methods were chosen to have different learning capacities.



From: https://scikit-learn.org/stable/modules/tree.html

We can study the impact of simulation complexity on ML models.

- Questions we can start to consider with a simulation testbed:
  - How does the complexity of the training environment impact learnability and generalizability of ML models?
  - Can synthetic data from simulations effectively supplement real world data?
  - Can synthetic data from simulations help assess how well ML models address concept drift?
  - Are some ML models better suited for highly complex scenarios?

How does the complexity of the environment impact the learnability and generalizability of ML models?