Learning **Dynamic Multimodal Networks**

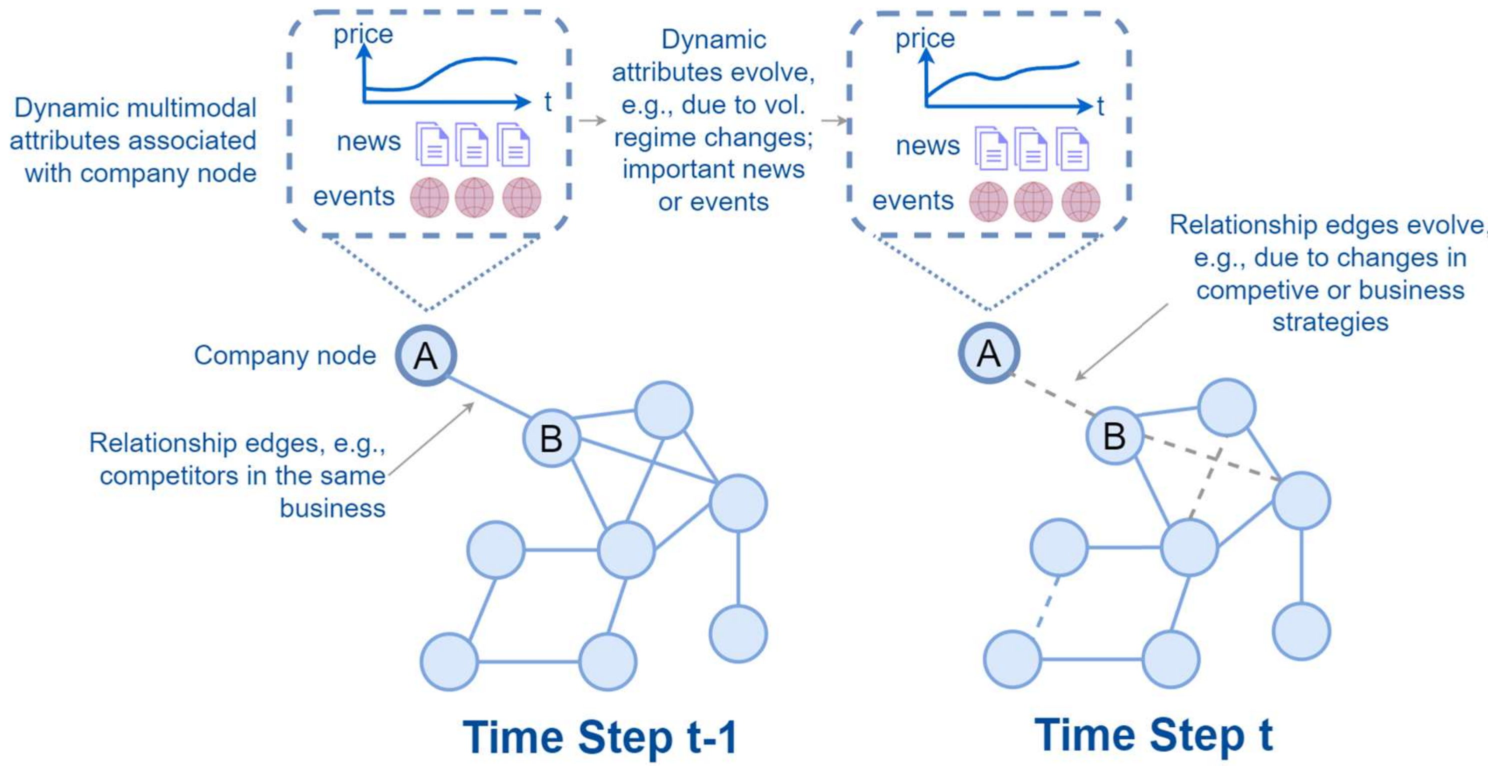
Gary Ang

Dissertation Defense, 22 May 2023

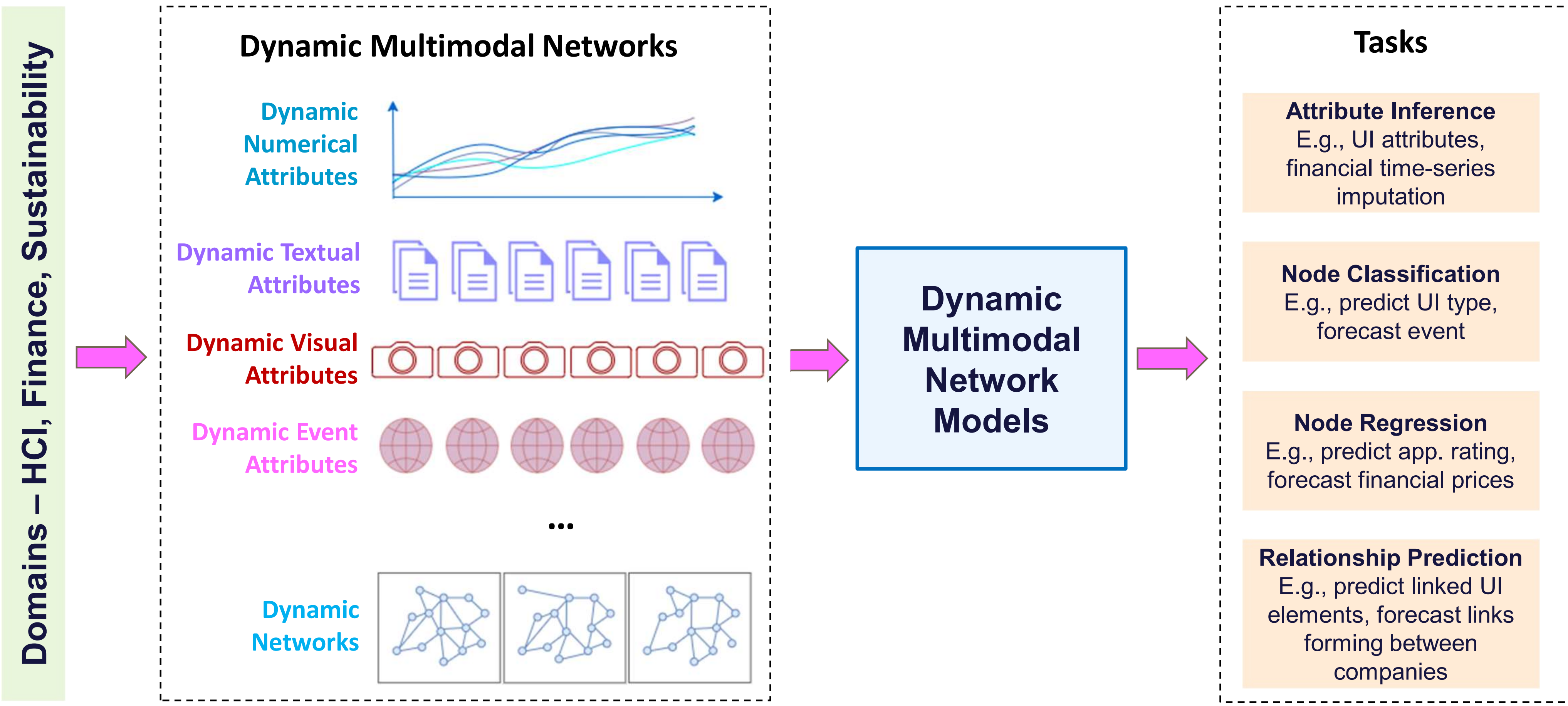
Presentation Outline

- Overview
- Related Work
- Research Summary
- Representative Works
- Conclusion & Future Research

Overview: Dynamic Multimodal Networks



Overview: Modeling of Dynamic Multimodal Networks



Overview: Research Objectives and Framework

Learning Dynamic Multimodal Networks

Progressively research on i) modelling multimodality and dynamicity in networks; and ii) interpreting role and contribution of multimodality and dynamicity of networks on predictive tasks

Multimodal Attributes + Static Networks

Part I

Dynamic Multimodal Attributes + Static Networks

Part II

Dynamic Multimodal Attributes + Dynamic Networks

Part III

Overview: Key Definitions

- **Dynamic networks**

$$G_t^c = [g^c(t-K^c), ..., g^c(t-1)]$$

  - $g^c(t-k) = (v^c(t-k), e^c(t-k))$ is the snapshot at time-step $(t-k)$ for network of type c

- **Dynamic multimodal node attributes** (from modality m, out of M modalities)

$$X^m_{V_t^c,t} = [x^m_{V_t^c}(t-K^m), ..., x^m_{V_t^c}(t-1)] \in \mathbb{R}^{|V_t^c| \times K^m \times d^m}$$

  - K^c and K^m are number of time-steps in windows of network and m^{th} attribute modality, respectively.

Overview: Types of Networks

I. Static Networks with **Multimodal** Attributes
$\boldsymbol{M} > \mathbf{1}, K^{\mathrm{m}} = \mathrm{K}^{\mathrm{c}} = 1$

II. Static Networks with **Dynamic Multimodal** Attributes
$M > 1, \boldsymbol{K^{m}} > \mathbf{1}, \mathrm{K}^{\mathrm{c}} = 1$

III. **Dynamic** Networks with **Dynamic Multimodal** Attributes
$M > 1, K^{\mathrm{m}} > 1, \boldsymbol{K^{\mathrm{c}}} > \mathbf{1}$

Related Works

- Most random walk models do not capture attributes
- Most GNN models assume attributes static and/or unimodal

- Most do not focus on attributes or only unimodal attributes

- Capture time-series attributes but unimodal, num.
- Static networks

- Usually do not capture networks
- Unimodal time-series attributes

Attribute			Network		Related Work
Static $K^m = 1$	Dynamic $K^m > 1$	Multimodal $M > 1$	Static $K^c = 1$	Dynamic $K^c > 1$	Modelling approaches and examples of key works.
●	○	◐	●	○	**Static Network Models**: DeepWalk [116], node2vec [52], metapath2vec, SDNE [150], GVAE [74], CAN [102], GCN [75], GraphSAGE [55], GAT [146], HGT [60]. Some static network models are designed for multimodal attributes, but only capture two modalities - visual and textual [126,157].
○	◐	○	○	●	**Dynamic Network Models**: DANE [80], EvolveGCN [112], VGRNN [53], DynGEM [50], TGAT [167]. Most dynamic network models assume time-stamps of dynamic attributes are aligned with time-stamps of dynamic networks, i.e. they cannot capture sequences of attributes and networks that are of different lengths or granularities.
○	●	○	●	○	**Spatio-Temporal Models**: DCRNN [84], STFGCN [177], GC-LSTM [24], StemGNN [37], MTGNN [164].
○	●	◐	◐	○	**Time-Series Models**: NBEATS [107], TST [179], TFT [89], DARNN [118], DyGAP [143]. Most time-series models capture unimodal numerical attributes. Some financial time-series models, such as VolTAGE [129], are designed for dynamic multimodal attributes and networks, but only capture two modalities, e.g., text and audio of calls, and static networks.

Note: ●: Works address this setting; ◐: Some works address this setting; ○: Works do not address this setting.

Summary of Research

Part I: Networks with *Multimodal* Attributes

Key Aspects

- Model networks with multimodal attributes
- Capture different types of positional attributes
- Model different types of nodes and relationships
- Interpret contribution of different networks and multimodal attributes

Proposed MAAN, EMAAN, HAMP, AHAMP models

Research Objective	Attribute			Network	
Work based on Dissertation (Model)	Static $K^m = 1$	Dynamic $K^m > 1$	Multimodal $M > 1$	Static $K^c = 1$	Dynamic $K^c > 1$
Networks with Multimodal Attributes	✓	-	✓	✓	-
Learning Network-Based Multi-Modal Mobile User Interface Embeddings, ACM IUI 2021 (MAAN)	MAAN utilizes attention mechanisms across multiple channels and stages to model bipartite network structural information and multimodal node attributes.				
Learning Semantically Rich Network-Based Multi-Modal Mobile User Interface Embeddings, ACM TiiS (EMAAN)	EMAAN extends MAAN to capture edge attributes with additive and multiplicative methods.				
Learning User Interface Semantics from Heterogeneous Networks with Multimodal and Positional Attributes, ACM IUI 2022 (HAMP)	HAMP utilizes a positional vectorizer and graph transformer to model heterogeneous networks with multimodal and positional node attributes.				
Learning and Understanding User Interface Semantics from Heterogeneous Networks with Multimodal and Positional Attributes, ACM TiiS (AHAMP)	AHAMP extends HAMP with an adaptive graph discovery and interpretability method to discover and understand important network edges.				

Summary of Research

Part II: Networks with *Dynamic* Multimodal Attributes

Key Aspects

- Model networks with dynamic multimodal attributes
- Address low signal-to-noise and non-stationarity of dynamic multimodal attributes
- Capture global (relevant to all network nodes), and local (specific to node) dynamic multimodal attributes

Proposed KECE, GLAM models

Research Objective	Attribute			Network	
Work based on Dissertation (Model)	Static $K^m = 1$	Dynamic $K^m > 1$	Multimodal $M > 1$	Static $K^c = 1$	Dynamic $K^c > 1$
Networks with Dynamic Multimodal Attributes	✓	✓	✓	✓	-
Learning Knowledge-Enriched Company Embeddings for Investment Management, ACM ICAIF 2021 (KECE)	KECE utilizes attention mechanisms to model knowledge graphs with multimodal attributes, and learn knowledge-enriched node embeddings.				
Investment and Risk Management with Online News and Heterogeneous Networks, ACM TWEB (GLAM)	GLAM models global and local information from multiple modalities and heterogeneous networks, and uses an adaptive curriculum learning method to isolate significant changes in time-series dynamics to address noisy time-series information.				

Summary of Research

Part III: Dynamic Networks with Dynamic Multimodal Attributes

Proposed GAME, DynMix, DynScan models

Key Aspects

- Model dynamic networks with dynamic multimodal attributes
- Address low signal-to-noise and non-stationarity of dynamic networks
- Capture evolving dependencies between nodes due to dynamic multimodal attributes, i.e., multiple types of dynamic implicit networks
- Interpret contribution of dynamic networks and dynamic multimodal attributes

Research Area	**Attribute**			**Network**	
Work based on Dissertation (Model)	Static $K^m = 1$	Dynamic $K^m > 1$	Multimodal $M > 1$	Static $K^c = 1$	Dynamic $K^c > 1$
Dynamic Networks with Dynamic Multimodal Attributes	✓	✓	✓	✓	✓
Guided Attention Multimodal Multitask Financial Forecasting with Inter-Company Relationships and Global and Local News, ACL 2022 (GAME)	GAME encodes local information of different lengths and frequencies, learns dynamic implicit network weights to update explicit network relationships, and uses cross attention between local and global multimodal information to guide learning of relevant global information.				
Learning Dynamic Multimodal Implicit and Explicit Networks for Multiple Financial Tasks, IEEE BigData 2022 (DynMix)	DynMix discovers dynamic implicit networks from multimodal time-series. The dynamic implicit networks are paired with dynamic explicit networks to form different views for a dynamic self-supervised learning approach to address noisy non-stationary time-series distributions and correlations.				
Learning Dynamic Multimodal Network Slot Concepts from the Web for Forecasting Environmental, Social and Governance Ratings, Under review at TWEB (DynScan)	DynScan learns dynamic multimodal network slot concepts with attention mechanisms, and inter-modality alignment and intra-modality disentanglement loss functions. The learnt slot concepts are utilized to address noisy and non-stationary dynamic multimodal network information.				

Representative Works

to be covered in this presentation

I. Human Computer Interaction (HCI) Domain

Heterogeneous Attention-based Multimodal Positional (HAMP) network model

- *Learning User Interface Semantics from Heterogeneous Networks with Multimodal and Positional Attributes, ACM IUI 2022, Honorable Mention*

Adaptive HAMP (AHAMP) network model

- *Learning and Understanding User Interface Semantics from Heterogeneous Networks with Multimodal and Positional Attributes, ACM TiiS*

II. Financial Domain

Dynamic Multimodal Multitask Implicit Explicit (DynMix) *network model*

- *Learning Dynamic Multimodal Implicit and Explicit Networks for Multiple Financial Tasks, IEEE BigData 2022*

III. Sustainability Domain

Dynamic Multimodal Slot Concept Attention-based Network (DynScan) *model*

- *Learning Dynamic Multimodal Network Slot Concepts from the Web for Forecasting Environmental, Social and Governance Ratings, ACM TWEB, Submitted & Under Review*

4 of the other works listed earlier were already shared during the dissertation proposal

Representative Work I
HAMP and AHAMP - Modeling Networks with Multimodal Attributes

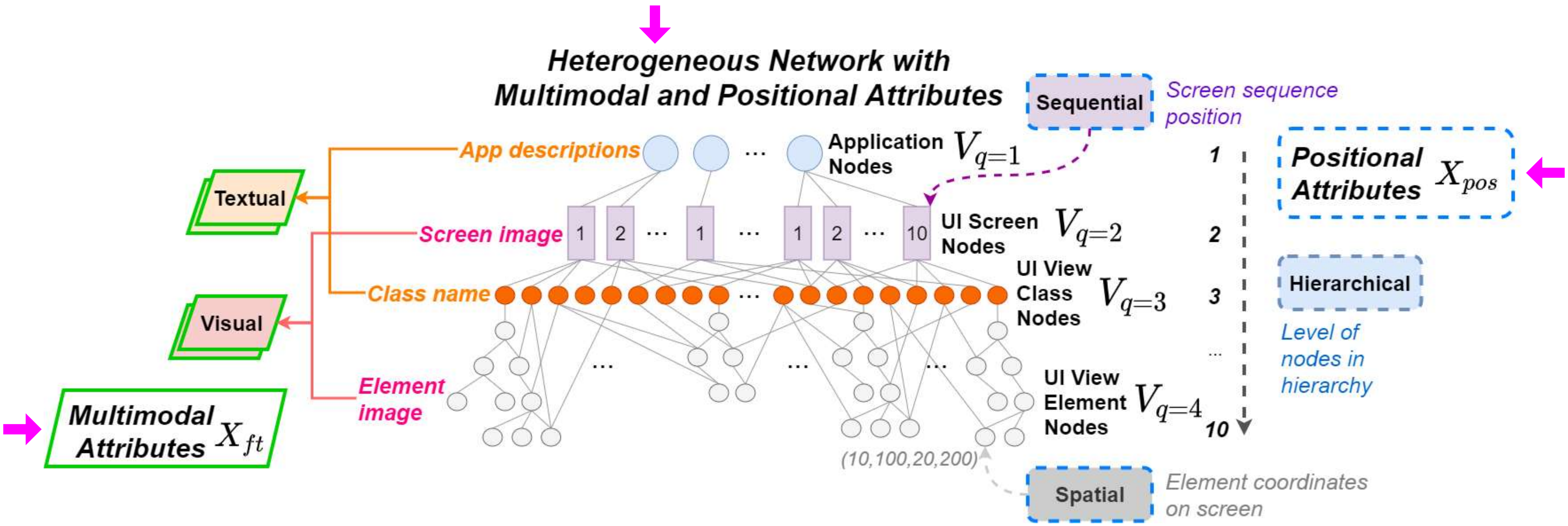
Heterogeneous Attention-based Multimodal Positional (HAMP) network model

- *Learning User Interface Semantics from Heterogeneous Networks with Multimodal and Positional Attributes, ACM IUI 2022, Honorable Mention*

Adaptive HAMP (AHAMP) network model

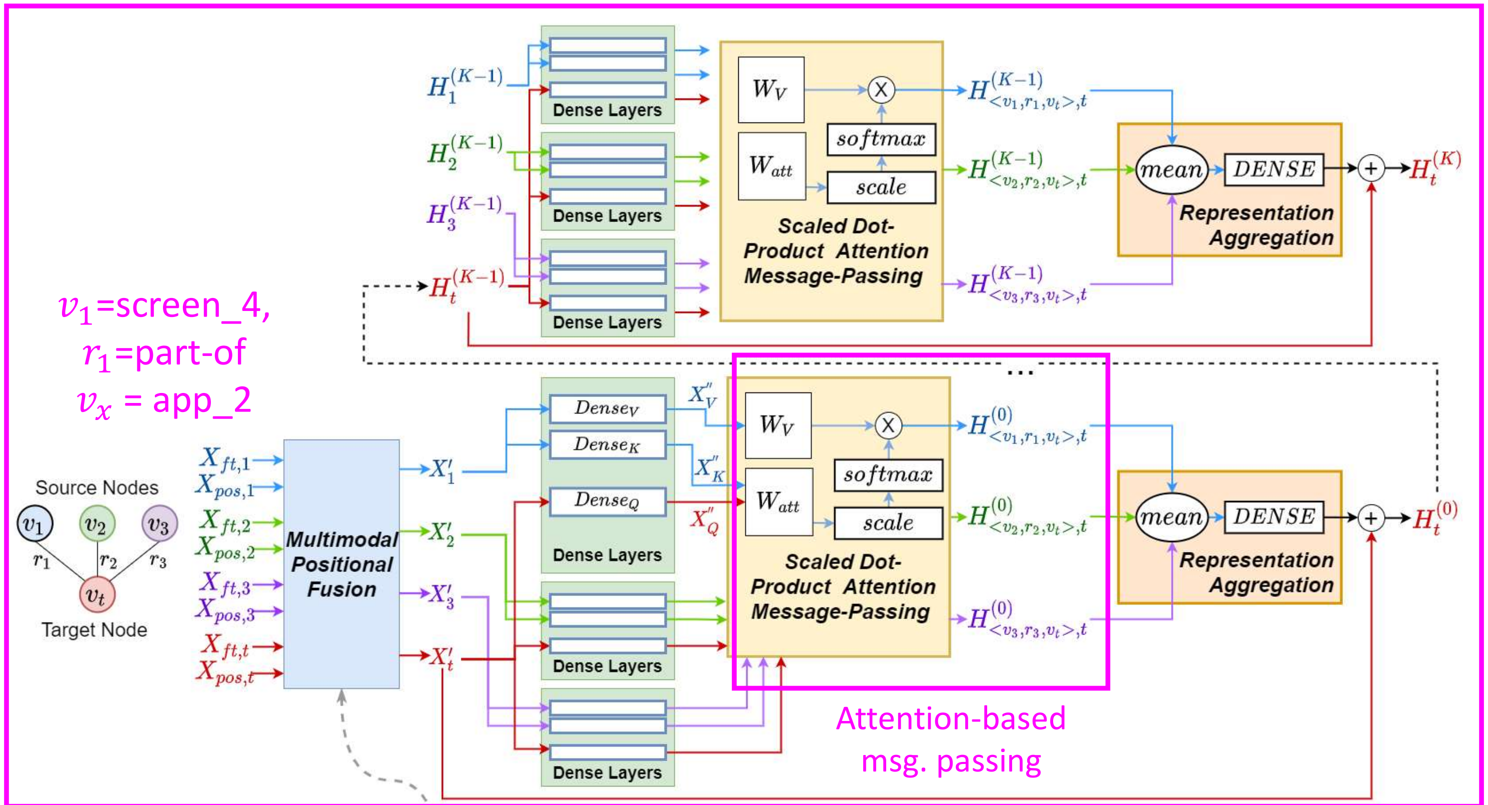
- *Learning and Understanding User Interface Semantics from Heterogeneous Networks with Multimodal and Positional Attributes, ACM TiiS*

Framing UIs as Networks

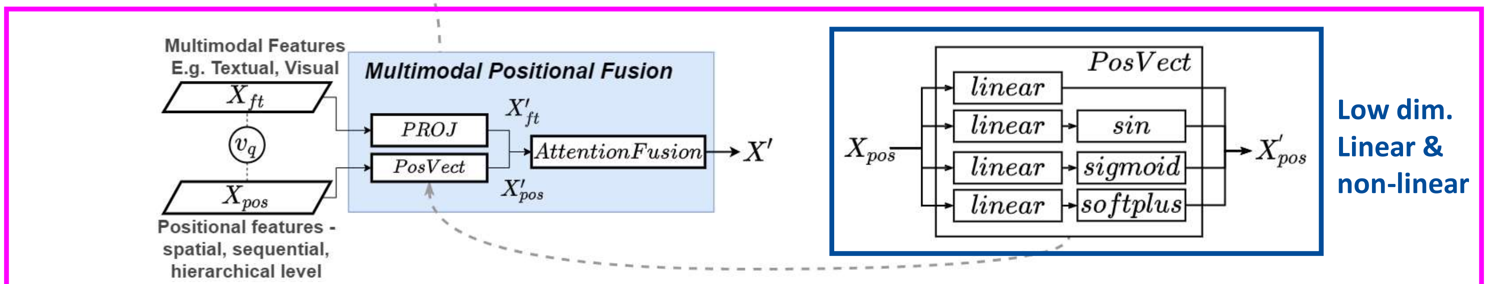


Overview

Modeling different node and edge types



Capturing multimodal & positional attributes



From *heterogeneous network message passing* to *adaptiveness* and *interpretability*

Heterogeneous network encoding in HAMP

$$AttScore_{\langle v_s,r,v_t\rangle} = softmax_{v_s \in N(v_t)} scale(X''_{K,v_s} W_{att} X''_{Q,v_t})$$

$$H_{\langle v_s,r,v_t\rangle,t} = \sum_{v_s \in N(v_t)} AttScore_{\langle v_s,r,v_t\rangle} \cdot X''_{V,v_s} W_V$$

Update node representations with learnt attention scores across multiple tuples

Then aggregate by taking mean

Adaptiveness and interpretability in AHAMP

Learnable mask

$$H^{(k)}_{\langle v_s,r,v_t\rangle,t} = \sum_{v_s \in N(v_t)} \boxed{\tilde{e}_{l,\langle v_s,r,v_t\rangle}} \odot (\mathrm{AttScore}^{(k)}_{\langle v_s,r,v_t\rangle} \cdot X''^{(k)}_{V,v_s} W^{(k)}_V)$$

$$\mathcal{L}' = \boxed{\mathcal{L}(y, \hat{y} = \mathrm{f}_{\theta,\mathrm{E_L}}(G_L = (V,X)))}$$

Prediction conditioned on masked graph

$$\boxed{-\alpha \sum_{\langle v_s,r,v_t\rangle \in E} \tilde{e}_{l,\langle v_s,r,v_t\rangle} \log(\tilde{e}_{l,\langle v_s,r,v_t\rangle}) + (1 - \tilde{e}_{l,\langle v_s,r,v_t\rangle}) \log(1 - \tilde{e}_{l,\langle v_s,r,v_t\rangle})}$$

Entropy regularization

$$\boxed{+\beta \sum_{\langle v_s,r,v_t\rangle \in E} \tilde{e}_{l,\langle v_s,r,v_t\rangle}}$$

Sparsity regularization

Datasets

- Extracted from RICO and ENRICO
- RICO largest mobile app. dataset; ENRICO is subset of RICO with manually annotated topic labels (*e.g., dialer, news, tutorial topics*)

Datasets	RICO-N	RICO-O	ENRICO
Num. of Application Nodes	1000	1000	869
Num. of UI Screen Nodes	5879	9108	1460
Num. of UI View Class Nodes	1563	2920	1506
Num. of UI Element Nodes	109,387	203,522	28,821
Num. of App - UI Screens Edges	5879	9108	1460
Num. of UI Screens - UI View Classes Edges	38,961	68,305	10,113
Num. of UI View Classes - UI Elements Edges	109,387	203,522	28,821
Length of longest sequence	36	46	38
Max. depth of hierarchy	9	10	9
Num. of UI element component-types	26	26	-
Num. of UI screen genres	36	33	-
Num. of UI screen topics	-	-	20
Range of mobile app. ratings	1.15 to 4.92	1.72 to 4.91	-

Selected Experiment Results

UI Screen Genre Classification

	RICO-N		RICO-O	
	Micro F1	**Macro F1**	**Micro F1**	**Macro F1**
Log. Regression	0.127	0.059	0.153	0.043
GCN	0.087	0.048	0.137	0.042
SGC	0.046	0.011	0.113	0.012
GraphSAGE	0.079	0.035	0.136	0.038
GAT	0.079	0.058	0.159	0.063
hGAO	0.087	0.067	0.168	0.060
HAN	0.698	0.648	0.517	0.298
Screen2Vec	0.392	0.311	0.466	0.407
HAMP	0.970	0.877	0.921	0.759
AHAMP	**0.993**	**0.966**	**0.997**	**0.962**

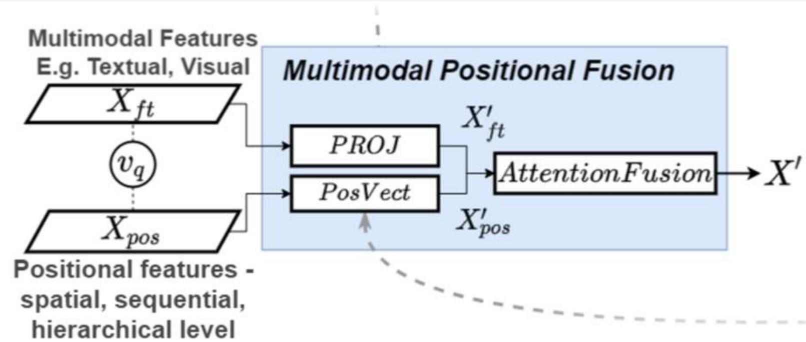
UI Screen Topic Classification

	ENRICO	
	Micro F1	**Macro F1**
Log. Regression	0.264	0.118
GCN	0.290	0.110
SGC	0.179	0.016
GraphSAGE	0.335	0.231
GAT	0.305	0.196
hGAO	0.390	0.278
HAN	0.452	0.436
Screen2Vec	0.336	0.206
HAMP	**0.996**	0.996
AHAMP	**0.996**	**0.997**

- HAMP and AHAMP outperform baselines across 4 tasks - **UI screen genre** and **topic classification**; **UI element component-type classification**; **app. rating regression**. Selected results above
    - In general, HAN (which models heterogeneous networks); and Screen2Vec(which models multimodal information) closest

Interpretability

Selected Results



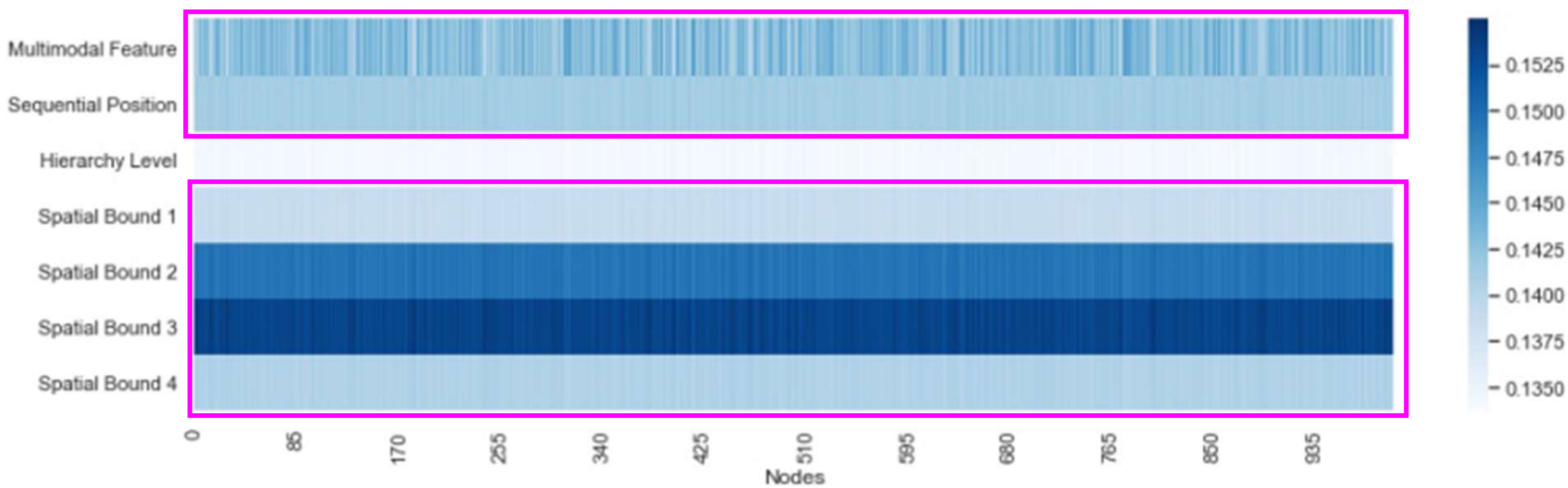


Fig. 8. Learnt Attention Weights for App. Rating Regression (RICO-N). Darker shade of blue indicates higher attention weight β_m for the node. We see that multimodal feature, sequential position and spatial bounds 2 and 3 are higher.

Interpretability

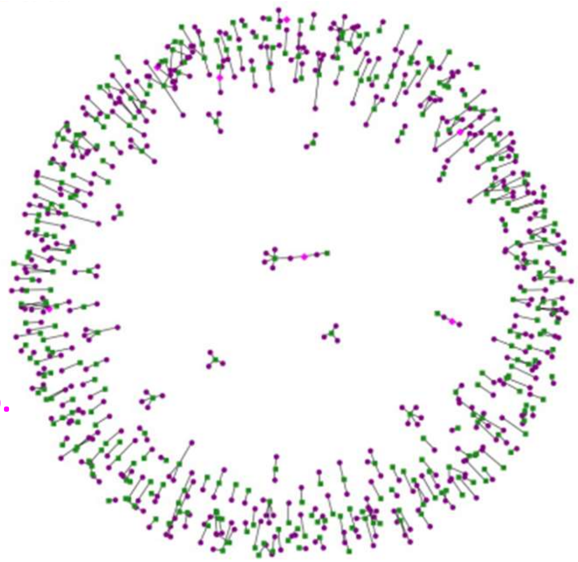
Selected Results

Top 500 edges from learnable mask

$\tilde{e}_{l,\langle v_s,r,v_t\rangle}$

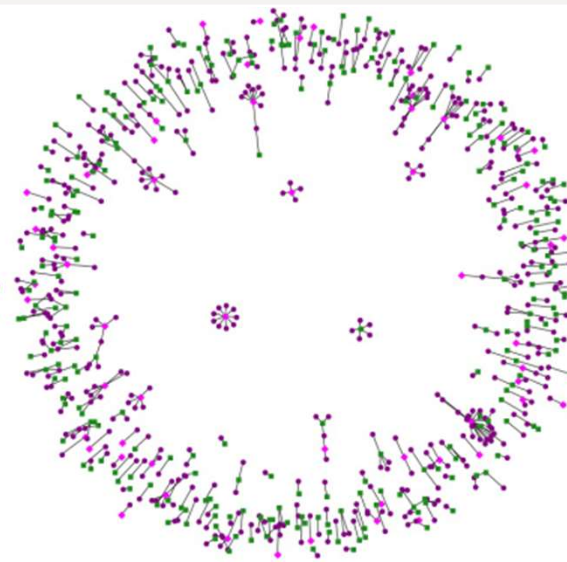
UI Screen Genre Classification

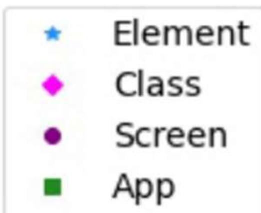
Screen genre depends on app.



2-hop App-Screen-Class r/s as app. rating depends on classes used

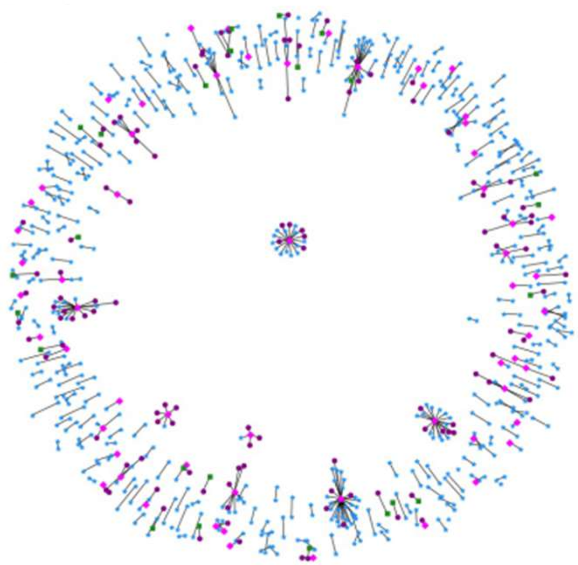
App. Rating Regression





UI Element Type Classification

2-hop Element-Class-Screen r/s as type depends on class and role in screen



3-hop r/s in a cluster as UI screen topic depends on information across multiple nodes

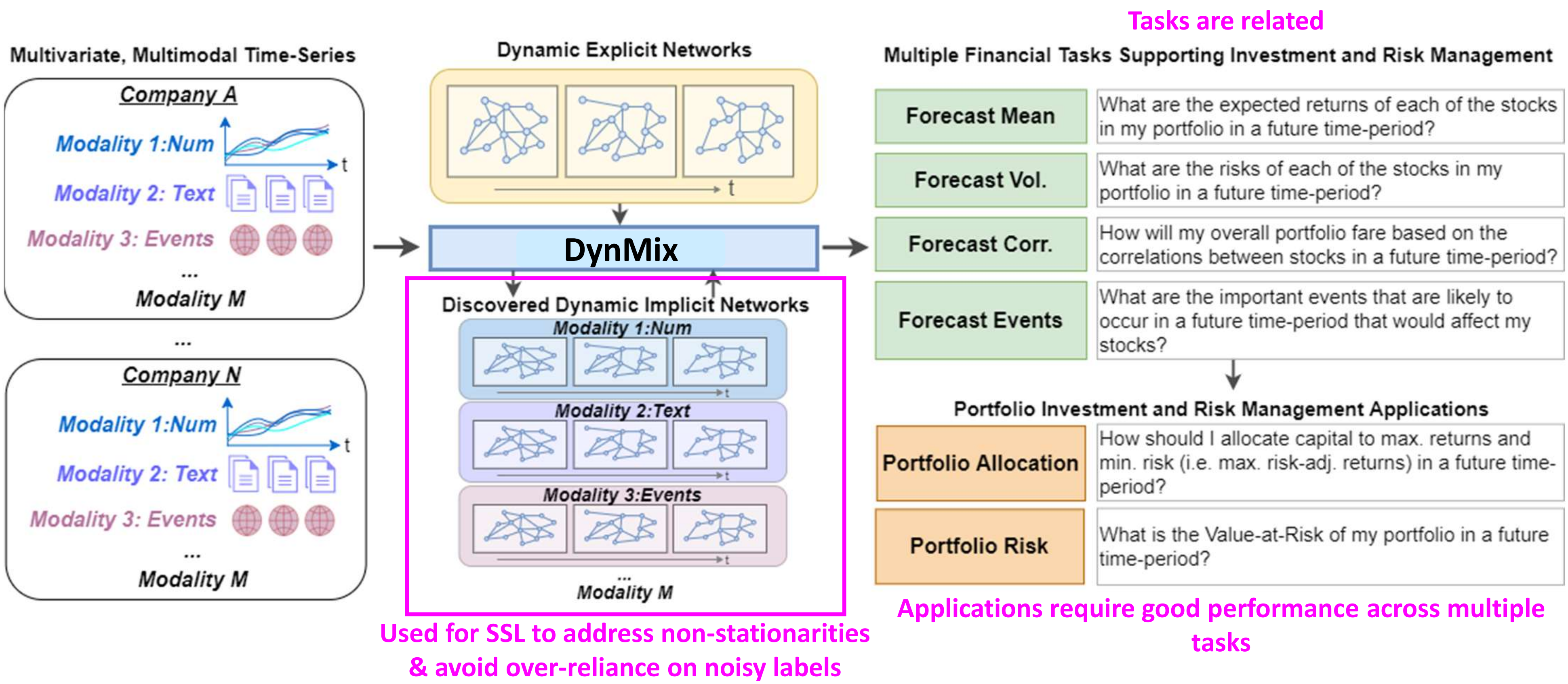
UI Screen Topic Classification

Representative Work II
DynMix - Modeling Dynamic Networks with Dynamic Multimodal Attributes

Dynamic Multimodal Multitask Implicit Explicit (DynMix) network model

- *Learning Dynamic Multimodal Implicit and Explicit Networks for Multiple Financial Tasks, IEEE BigData 2022*

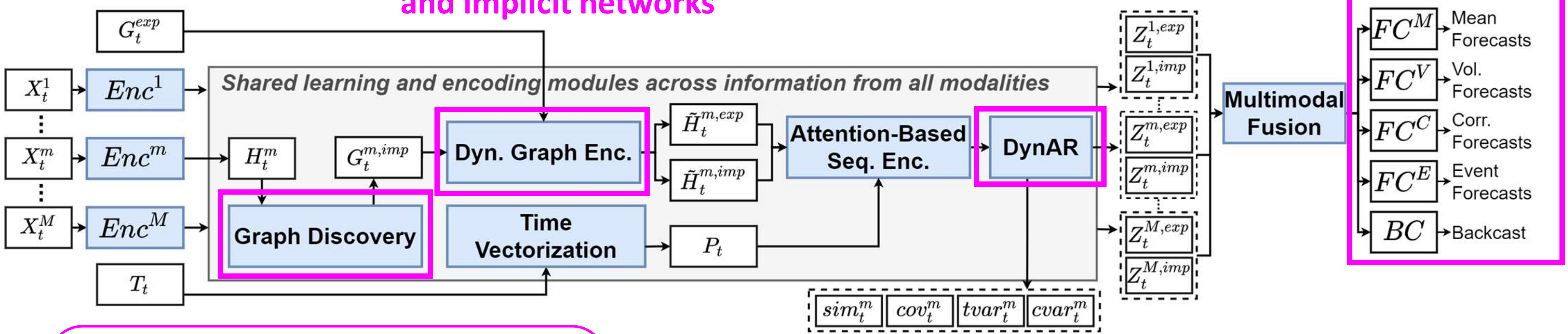
Capturing *dynamic networks* and *dynamic multimodal information* for *multiple financial tasks*



Overview

Dynamic graph encoding of explicit and implicit networks

Multitask learning

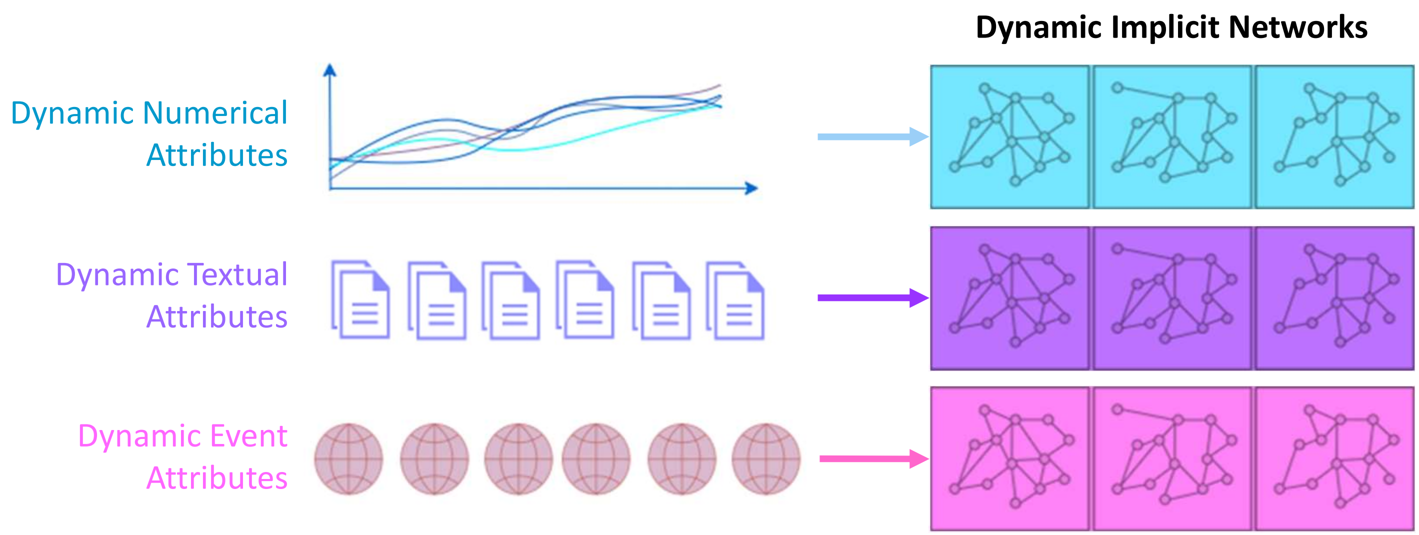


Discovering dynamic networks from dynamic information from multiple modalities to model evolving dependencies between nodes

Dynamic alignment and regularization with self-supervised learning

Using Dynamic Implicit Networks

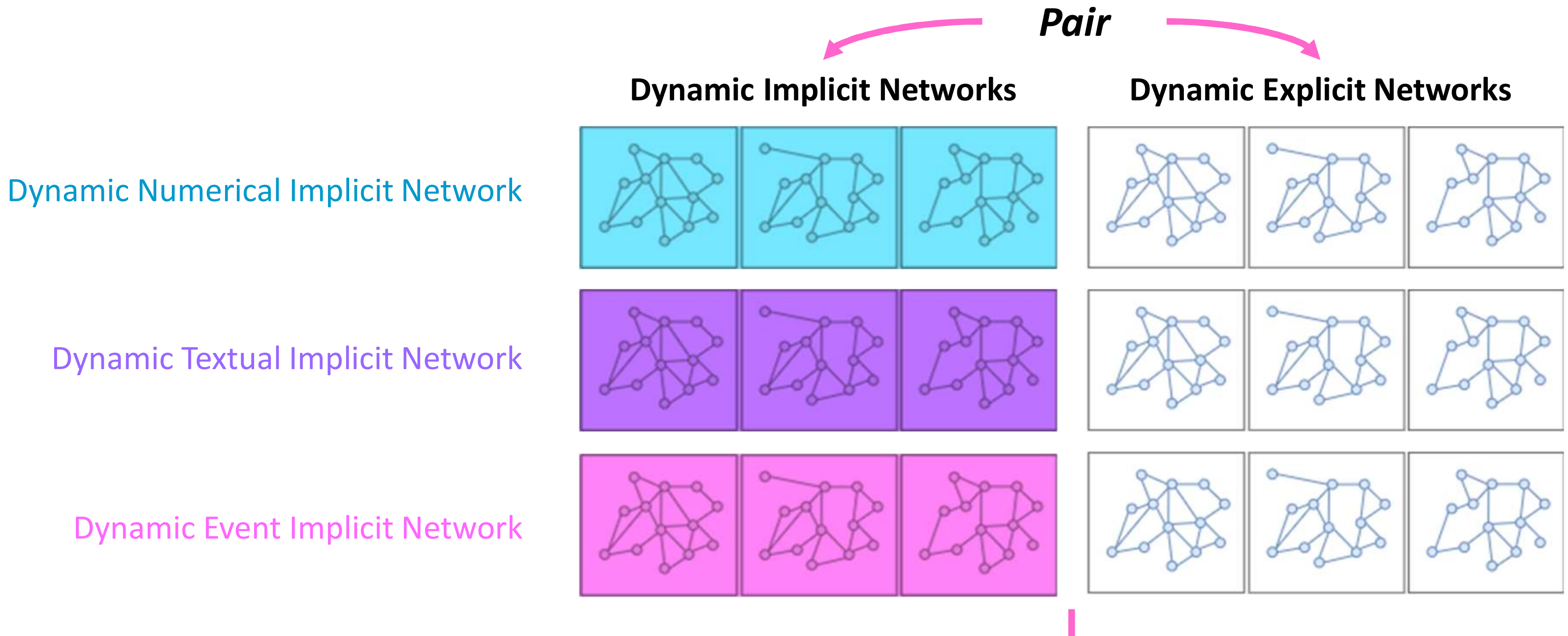
Capture evolving dependencies between nodes due to dynamic multimodal attributes



Learn dynamic underlying networks for information from different modalities

For Dynamic Self-Supervised Learning

Align and regularize dynamic networks and time series from different modalities to extract dynamic signals

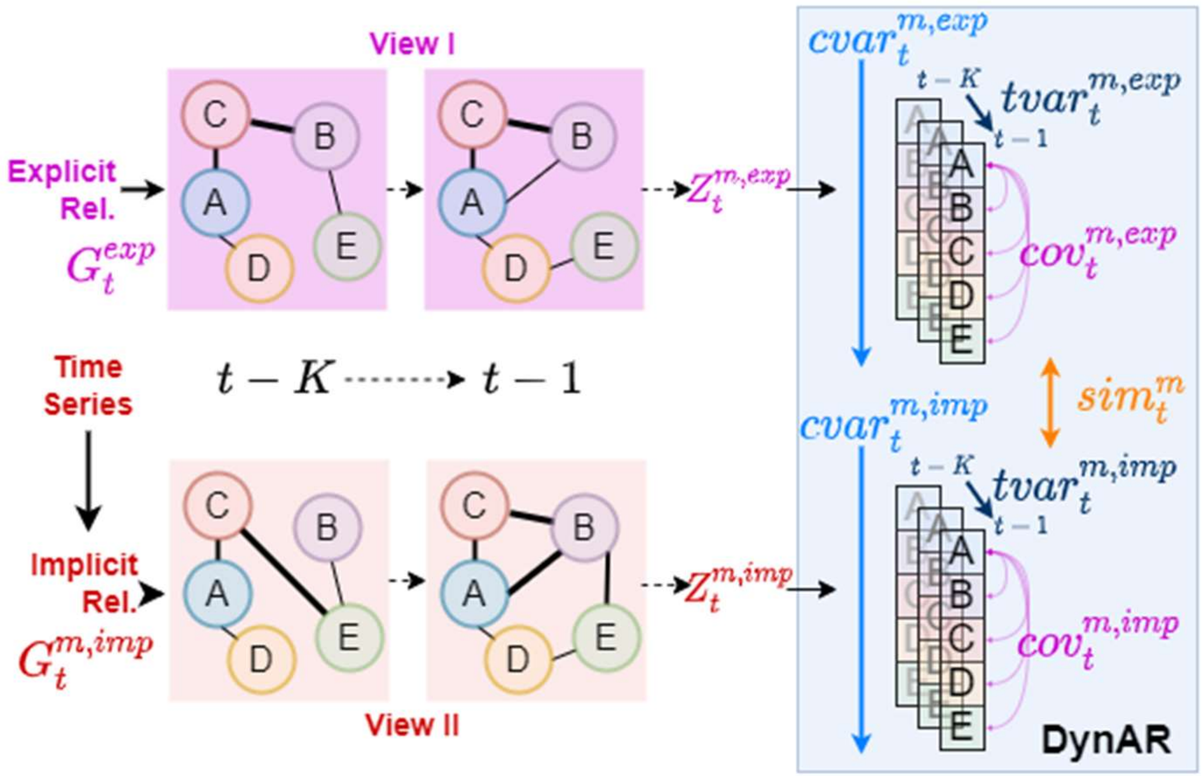


No need for augmentation to generate 2nd view!

Dynamic Self Supervised Learning

For Dynamic Self-Supervised Learning

Dynamic *alignment* and regularization with self-supervised learning

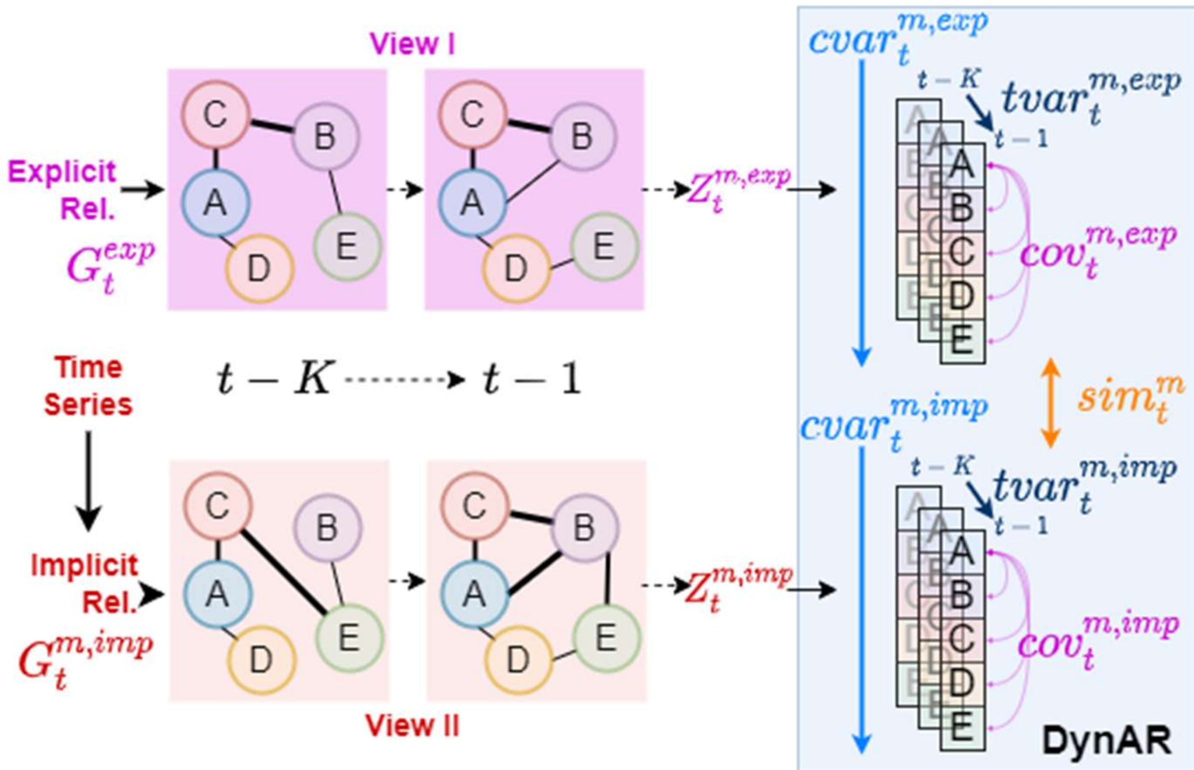


sim_t^m - minimize distance between representations of same companies

$cov_t^{m,exp/imp}$ - de-correlates representations of different companies

For Dynamic Self-Supervised Learning

Dynamic alignment and *regularization* with self-supervised learning



$cvar_t^{m,exp/imp}$ - regularization with dynamic level of variance across **companies**

Maintain min. dynamic level of variation of representations across companies and time to prevent mode collapse

$tvar_t^{m,exp/imp}$ - regularization with dynamic level of variance across **time**

Datasets

- Dynamic networks: Global Database of Events, Language and Tone (GDELT) Global Knowledge Graphs (GKG)
- Dynamic multimodal attributes: Time series of stock price, news articles, and events

	IN-NY	IN-NA	BE-NY	BE-NA
Num. articles	189,917		1,295,491	
Num. stocks	336	371	1,693	1,705
Num. edges	2,212±1,574	1,217±387	4,913±2,887	3,973±1,749
Avg. edge weights	3,547±2,911	3,901±2,218	1,537±1,386	1,541±1,448
Num. event types	2,177	2,135	2,182	2,186
Event density	0.45±0.16	0.33±0.13	0.09±0.03	0.09±0.03

Experiments

Forecast means, volatility, correlations

	IN-NY			IN-NA			BE-NY			BE-NA		
	RMSE	MAE	SMAPE	RMSE	MAE	SMAPE	RMSE	MAE	SMAPE	RMSE	MAE	SMAPE
Mean Forecasts												
GRU	0.0744	0.0144	1.4579	0.0356	0.0175	1.4801	0.1341	0.0343	1.4357	0.6119	0.0851	1.4587
TST	0.0742	0.0140	1.3844	0.0335	0.0155	1.3631	0.1286	0.0251	1.3139	0.5373	0.0637	1.5380
FAST	0.0742	0.0141	1.3511	0.0362	0.0164	1.7424	0.1464	0.0260	1.3403	0.6358	0.0673	1.3801
MTGNN	0.0712	0.0139	1.3002	0.0314	0.0149	1.4843	0.1386	0.0338	1.5168	0.4844	0.0609	1.3218
DYGAP	0.0723	0.0146	1.4430	0.0353	0.0157	1.4074	0.1430	0.0263	1.4411	0.6361	0.0670	1.3520
EVOLVEGCN-H	0.0743	0.0143	1.5058	0.0394	0.0176	1.3273	0.1470	0.0288	1.3540	0.6444	0.0823	1.4600
EVOLVEGCN-O	0.0750	0.0153	1.4372	0.0354	0.0158	1.2921	0.1453	0.0268	1.3679	0.6325	0.0726	1.4014
DYNMIX	**0.0539**	**0.0111**	**1.1693**	**0.0244**	**0.0128**	**1.2080**	**0.1058**	**0.0194**	**1.2391**	**0.4164**	**0.0471**	**1.2337**
Volatility Forecasts												
GRU	0.2331	0.0507	0.6244	0.1188	0.0599	0.6384	0.4181	0.1049	1.1047	1.9152	0.2418	0.9816
TST	0.2330	0.0483	0.5578	0.1109	0.0559	**0.6046**	0.3974	0.0894	0.7063	1.7047	0.2133	0.8463
FAST	0.2332	0.0485	0.5595	0.1244	0.0605	0.6338	0.4688	0.0990	0.7521	1.9993	0.2316	0.7935
MTGNN	0.2011	0.0529	0.6206	0.1077	0.0565	0.6178	0.4122	0.0919	0.7488	1.5066	0.2023	0.7860
DYGAP	0.2224	0.0497	**0.5475**	0.1234	0.0606	0.6359	0.4542	0.0928	0.7028	2.0003	0.2410	0.8589
EVOLVEGCN-H	0.2331	0.0485	0.5602	0.1253	0.0619	0.6470	0.4688	0.1012	0.7670	2.0097	0.2547	0.8832
EVOLVEGCN-O	0.2328	0.0491	0.5831	0.1246	0.0607	0.6370	0.4656	0.1010	0.7646	1.9970	0.2419	0.8616
DYNMIX	**0.1491**	**0.0410**	0.5483	**0.0870**	**0.0510**	0.6086	**0.3284**	**0.0763**	**0.7011**	**1.2956**	**0.1667**	**0.7227**
Correlation Forecasts												
GRU	0.5361	0.4671	1.5484	0.5163	0.4425	1.4072	0.5341	0.4600	1.4455	0.5180	0.4416	1.4193
TST	0.5145	0.4498	1.4646	0.5069	0.4400	1.4802	0.5235	0.4540	1.4553	0.5099	0.4374	1.4661
FAST	0.5087	0.4414	1.3395	0.5099	0.4394	1.4239	0.5246	0.4546	1.4533	0.5112	0.4380	1.4374
MTGNN	0.5215	0.4558	1.5012	0.5167	0.4438	1.4078	0.5288	0.4576	1.4505	0.5268	0.4398	1.5195
DYGAP	0.5415	0.4562	1.4074	0.5074	0.4362	1.4192	0.5125	0.4443	1.4460	0.5060	0.4346	1.4738
EVOLVEGCN-H	0.5104	0.4448	1.3996	0.5058	0.4390	1.4681	0.5192	0.4520	1.4842	0.5270	0.4426	1.5217
EVOLVEGCN-O	0.5118	0.4466	1.4194	0.5064	0.4390	1.4564	0.5228	0.4536	1.4592	0.5033	0.4348	1.5197
DYNMIX	**0.4272**	**0.3551**	**1.0844**	**0.4515**	**0.3810**	**1.2256**	**0.4898**	**0.4173**	**1.3042**	**0.4848**	**0.4095**	**1.3461**

- **Means:** Diff. between DynMIX and baselines clearer for larger BE datasets
- **Volatilities, correlations, events**: Harder tasks. Diff. between DynMIX and baselines even clearer

Forecast events (NDCG)

	IN-NY	IN-NA	BE-NY	BE-NA
GRU	0.5909	0.3336	0.3091	0.2336
TST	0.5705	0.3513	0.2870	0.2179
FAST	0.6106	0.3549	0.3074	0.2246
MTGNN	0.5907	**0.3710**	0.2911	0.2317
DYGAP	0.6113	0.3485	0.3069	0.2126
EVOLVEGCN-H	0.4757	0.2743	0.2136	0.1542
EVOLVEGCN-O	0.5717	0.3512	0.2904	0.2193
DYNMIX	**0.6182**	0.3626	**0.3163**	**0.2474**

Ablation Study

	RMSE	MAE	SMAPE
	Mean Forecasts		
w/o. dynamic SSL	0.0579	0.0124	1.1719
w/o dynamic var. target	0.0569	0.0121	1.1782
R=50%	0.0558	0.0121	1.1789
w/o. time vect.	0.0562	0.0121	1.1759
DynMIX	**0.0539**	**0.0111**	**1.1693**
	Volatility Forecasts		
w/o. dynamic SSL	0.1527	0.0422	0.5544
w/o dynamic var. target	0.1511	0.0420	0.5532
R=50%	0.1501	0.0415	0.5503
w/o. time vect.	0.1495	0.0419	0.5586
DynMIX	**0.1491**	**0.0410**	**0.5483**
	Correlation Forecasts		
w/o. dynamic SSL	0.4973	0.4302	1.3785
w/o dynamic var. target	0.4399	0.3685	1.1225
R=50%	0.4304	0.3590	1.0983
w/o. time vect.	0.4280	0.3562	1.0948
DynMIX	**0.4272**	**0.3551**	**1.0844**

	NDCG
	Event Forecasts
w/o. dynamic SSL	0.6059
w/o dynamic var. target	0.5617
R=50%	0.5565
w/o. time vect.	0.5861
DynMIX	**0.6182**

- Greatest drop in performance when
  - No dynamic SSL, or
  - No dynamic variance target (recall $cvar$ and $tvar$)

Applications

Portfolio Allocation and Value-at-Risk

APP. HIGHER BETTER FOR $\mathcal{R}$; LOWER BETTER FOR % BR.

	IN-NY		IN-NA	
	$\mathcal{R}$	% Br.	$\mathcal{R}$	% Br.
GRU	0.04	34.4%	1.20	2.2%
TST	0.05	22.8%	0.32	2.2%
FAST	0.03	15.6%	0.55	2.8%
MTGNN	0.67	4.4%	0.33	8.9%
DYGAP	0.41	6.7%	0.61	3.9%
EVOLVEGCN-H	0.01	13.9%	0.06	13.9%
EVOLVEGCN-O	0.06	15.6%	0.32	5.0%
DYNMIX	**2.47**	**3.9%**	**3.31**	**1.1%**

Portfolio Allocation Optimization

- Forecasts as inputs into mean-variance risk min.

$$max_{\mathbb{W}}\ (\mathbb{W}^T\mu - \lambda\mathbb{W}^T\Sigma\mathbb{W})$$

- Forecast of **means as $\boldsymbol{\mu}$**, forecast of **volatilities and correlations used to get $\boldsymbol{\Sigma}$**; λ is degree of risk aversion
- Use **forecasts to get optimal allocation weights** $\mathbb{W}$
- $E^{forecast}$ realized return of portfolio with forecast models; E^{naive} realized return of portfolio with historical mean and co-variance

$$\mathcal{R} = E^{forecast}/E^{naive}$$

Value-at-Risk (Portfolio Risk)

- **10 day 95% VaR of $1m - means 5% probability of losses exceeding $1m over 10 day horizon for portfolio of assets**
- VaR from forecasted volatilities and correlations
- Breach if actual realized loss > forecast VaR
- Percentage VaR breaches (% Br.), i.e. percentage of losses in testing set that led to VaR breaches

Representative Work III
DynScan - Modeling Dynamic Networks with Dynamic Multimodal Attributes

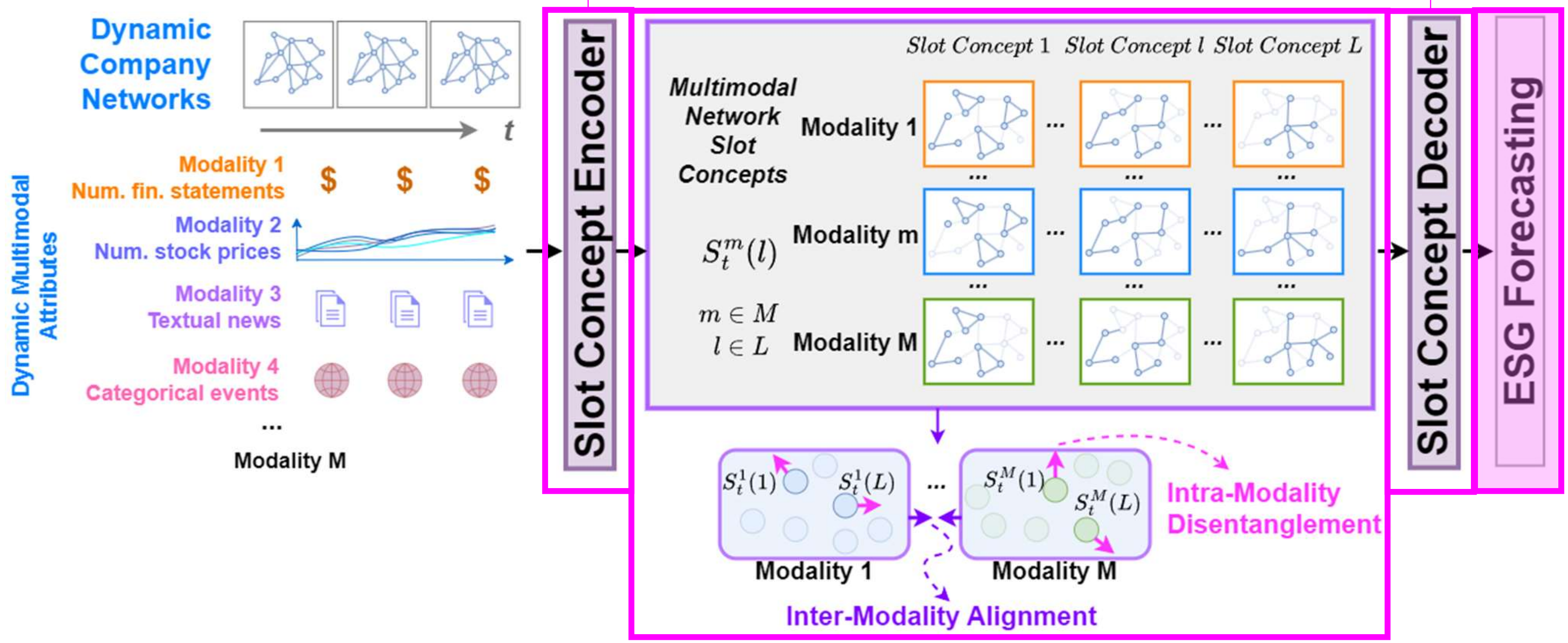
Dynamic Multimodal Slot Concept Attention-based Network (DynScan) model

- *Learning Dynamic Multimodal Network Slot Concepts from the Web for Forecasting Environmental, Social and Governance Ratings, ACM TWEB, Submitted & Under Review*

Overview

Attention-based slot concept encoder & decoder

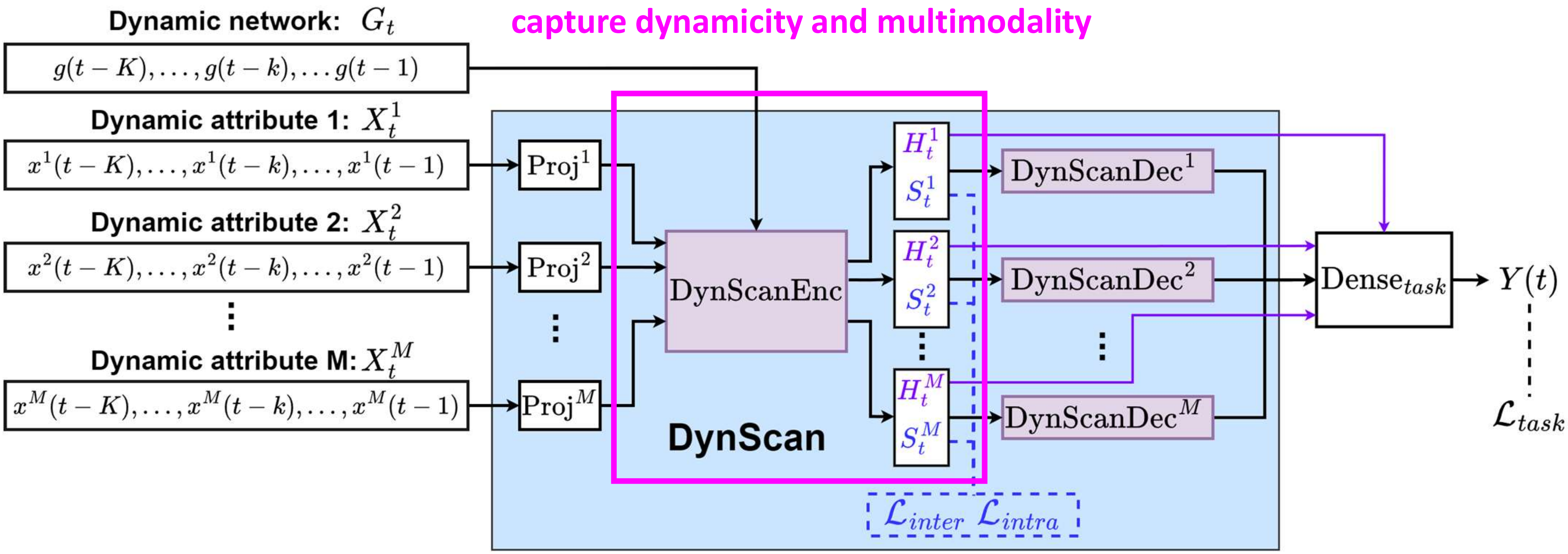
Inspired by how humans scan for high level relational concepts as they learn over time



Concepts more resistant to noise and distributional shifts over time

Overview

Learn underlying network slot concepts that capture dynamicity and multimodality

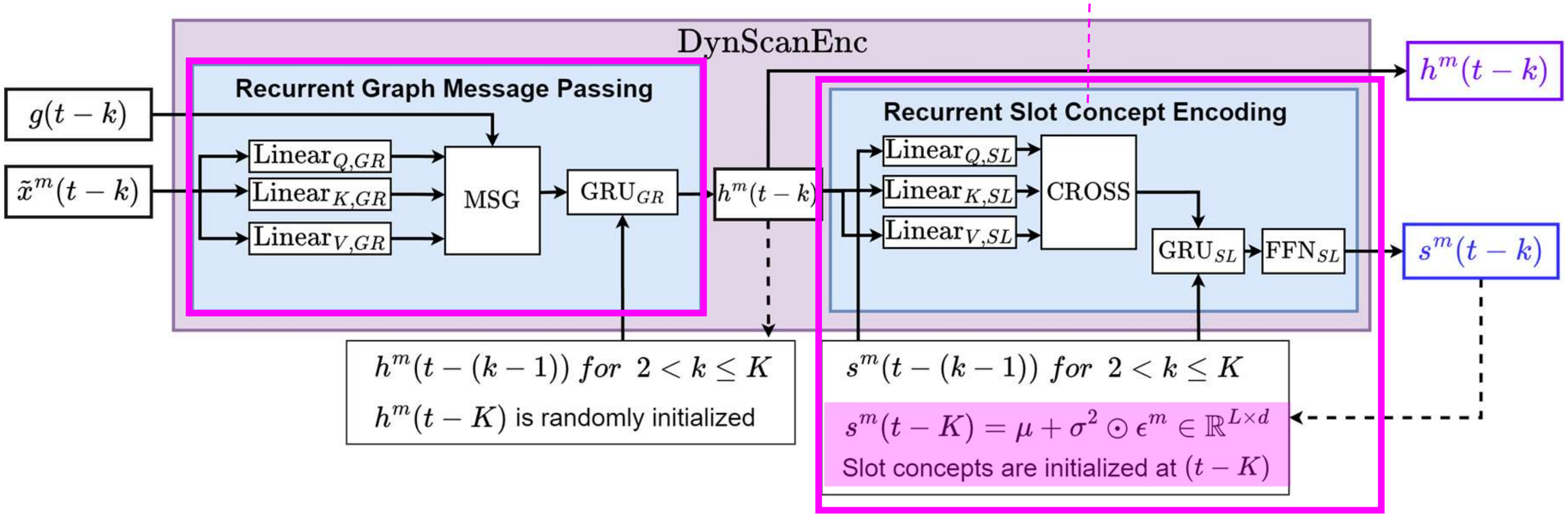


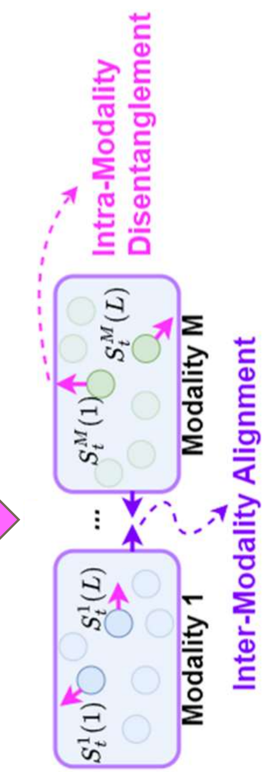
Dynamic Multimodal Slot Concept Learning

Sequential encoding of both dynamic networks and dynamic attributes

β_{SL}^{m}

Learn attention scores that represent affinity of each node to each of the slot concepts





Learn L slot concepts representing underlying dynamic multimodal network structures important for different prediction tasks

L ($\ll |V|$) slot concepts represent underlying subgraphs

Dynamic Multimodal Slot Concept Learning

Inter-modality slot concept alignment

Maximize first and second order similarities of distributions – same slot across modalities should be similar

$$\mathcal{L}_{inter} = \sum_{(m,m')\in \mathrm{P_m}} \left(\frac{1}{|b-a|} ||\mathbb{E}(S_t^m[t-1]) - \mathbb{E}(S_t^{m'}[t-1])||_2 + \frac{1}{|b-a|^2} ||\mathrm{C}(S_t^m[t-1]) - \mathrm{C}(S_t^{m'}[t-1])||_2 \right)$$

where (m, m') are the modality pairs and $C(x) = \mathbb{E}((x - \mathbb{E}(\mathrm{x}))^2)$

Intra-modality slot concept disentanglement

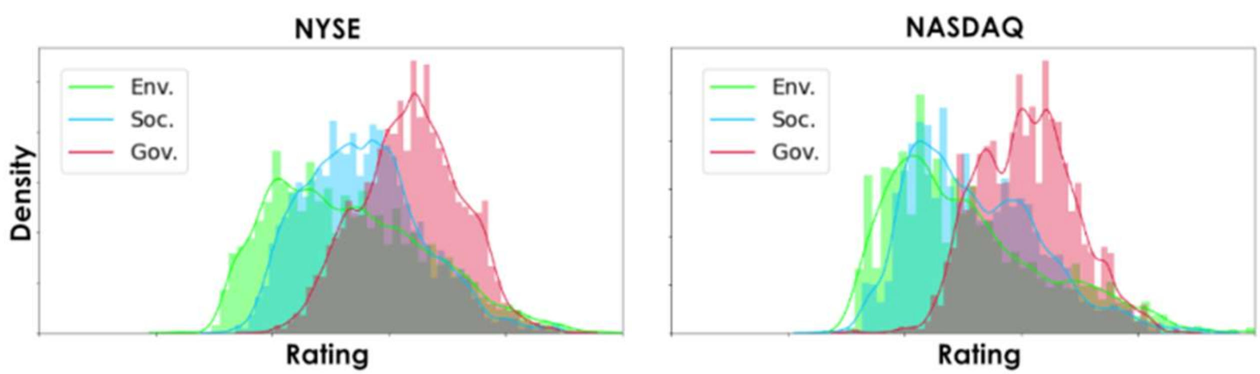
Minimize off-diagonals of covariances – different slots in one modality should be different

$$\mathcal{L}_{intra} = \mathrm{scale}\left(\sum_{m\in[1,M]} \sum_{v_i \neq v_j} [COV^m]^2_{v_i,v_j} \right)$$

where $COV^m = \frac{1}{|L|-1} \sum_{l\in L} (s^m[t-1](l) - \bar{s}^m[t-1])(s^m[t-1](l) - \bar{s}^m[t-1])^\top$

Datasets

- Dynamic networks: Global Database of Events, Language and Tone (GDELT) Global Knowledge Graphs (GKG)
- Dynamic multimodal attributes: Time series of stock price, news articles, and events
- **ESG ratings from Sustainalytics of NYSE (NY) and NASDAQ (NA) companies**



ESG Rating Distributions for NYSE (NY) and NASDAQ (NA) Companies.

Experiments

Selected Results

NY (NYSE)

	Env.			Soc.			Gov.		
	RMSE	MAE	SMAPE	RMSE	MAE	SMAPE	RMSE	MAE	SMAPE
XGBoost	15.548	12.311	23.1%	11.628	9.191	16.9%	10.331	8.292	13.2%
TST	15.023	11.666	22.8%	13.331	10.511	20.4%	14.512	12.223	21.0%
FAST	13.316	10.954	21.0%	10.514	8.626	16.1%	10.385	8.807	13.1%
GRU-GATv2	13.033	10.547	20.0%	9.826	7.801	14.4%	8.714	7.022	11.4%
EvolveGCN-H	22.945	17.512	39.5%	22.764	16.799	37.9%	25.048	18.142	37.1%
EvolveGCN-O	23.207	17.672	40.4%	22.822	16.827	38.9%	25.586	18.251	37.1%
DySAT	13.299	10.575	20.0%	10.004	7.937	14.6%	8.813	7.193	11.7%
DynMix	12.769	10.426	19.7%	9.583	7.739	14.3%	8.442	6.864	11.2%
DynScan	**11.733**	**9.463**	**18.0%**	**8.989**	**7.182**	**13.2%**	**8.136**	**6.618**	**10.7%**

- Difference between DynScan and baselines larger for Env./Soc. ratings
- Env. and Soc. ratings forecasting tasks generally harder
  - Gov. issues studied for a longer period, may be more stable and predictable
- Network information useful, but how it is captured important
  - DynMix closest. Among other baselines, GRU-GATv2, DySAT perform well, but not EvolveGCN
- Similar observations for companies on NASDAQ

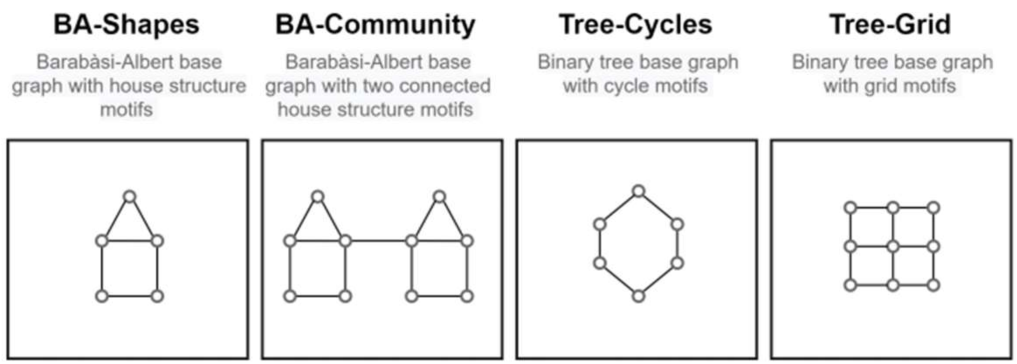
Ablation Study

	NY			NA		
	RMSE	MAE	SMAPE	RMSE	MAE	SMAPE
	Env.					
w/o. graph encoding	13.614	10.640	20.2%	12.476	9.732	19.1%
w/o. $\mathcal{L}_{inter}$	11.866	9.719	18.5%	11.382	8.904	17.5%
w/o. $\mathcal{L}_{intra}$	11.747	9.546	18.2%	11.314	<u>8.784</u>	17.3%
L=3	11.761	9.466	18.0%	**11.209**	8.786	<u>17.2%</u>
L=20	**11.727**	<u>9.414</u>	<u>17.9%</u>	11.291	8.850	17.5%
L=50	11.740	**9.352**	**17.8%**	11.254	8.804	<u>17.2%</u>
DynScan	<u>11.733</u>	9.463	18.0%	<u>11.239</u>	**8.668**	**17.1%**
	Soc.					
w/o. graph encoding	9.871	7.856	14.5%	10.184	8.097	15.8%
w/o. $\mathcal{L}_{inter}$	<u>9.025</u>	<u>7.196</u>	<u>13.3%</u>	9.574	7.614	<u>14.8%</u>
w/o. $\mathcal{L}_{intra}$	9.150	7.316	13.5%	9.584	7.634	14.9%
L=3	9.070	7.262	13.4%	9.554	7.582	<u>14.8%</u>
L=20	9.153	7.349	13.6%	**9.399**	<u>7.494</u>	**14.6%**
L=50	9.064	7.300	13.4%	<u>9.498</u>	7.596	<u>14.8%</u>
DynScan	**8.989**	**7.182**	**13.2%**	9.536	**7.484**	**14.6%**
	Gov.					
w/o. graph encoding	9.592	7.713	12.6%	9.033	7.318	12.3%
w/o. $\mathcal{L}_{inter}$	8.191	6.646	10.8%	8.032	6.672	11.3%
w/o. $\mathcal{L}_{intra}$	8.332	6.815	11.1%	7.961	6.580	11.1%
L=3	**8.081**	**6.565**	**10.6%**	<u>7.921</u>	6.523	<u>11.0%</u>
L=20	8.142	<u>6.614</u>	<u>10.7%</u>	7.943	<u>6.492</u>	<u>11.0%</u>
L=50	8.409	6.823	11.0%	**7.831**	6.539	11.1%
DynScan	<u>8.136</u>	6.618	<u>10.7%</u>	**7.831**	**6.451**	**10.9%**

- Graph encoding important
- Slot concept alignment and disentanglement important
- Performance varies for different number of slot concepts

Interpretability

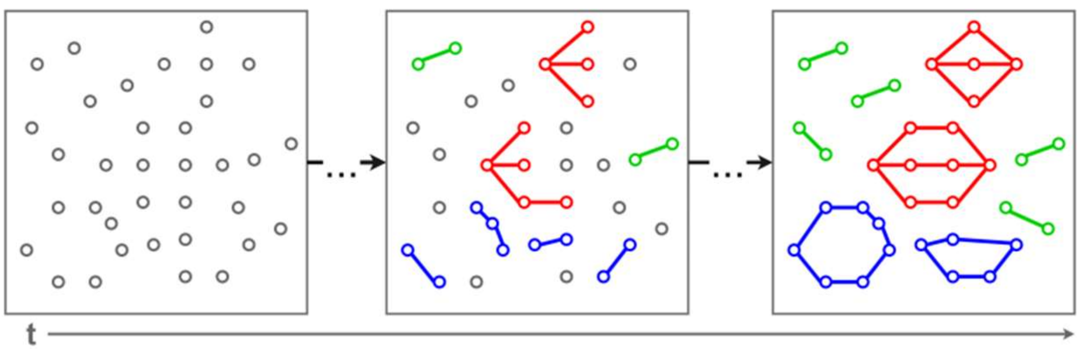
Synthetic Static Network Datasets



E.g., for BA-Shapes there are 4 motifs – part of

- base graph (label 0)
- middle of house (label 1)
- bottom of house (label 2)
- top of house (label 3)

Synthetic Dynamic Network Dataset

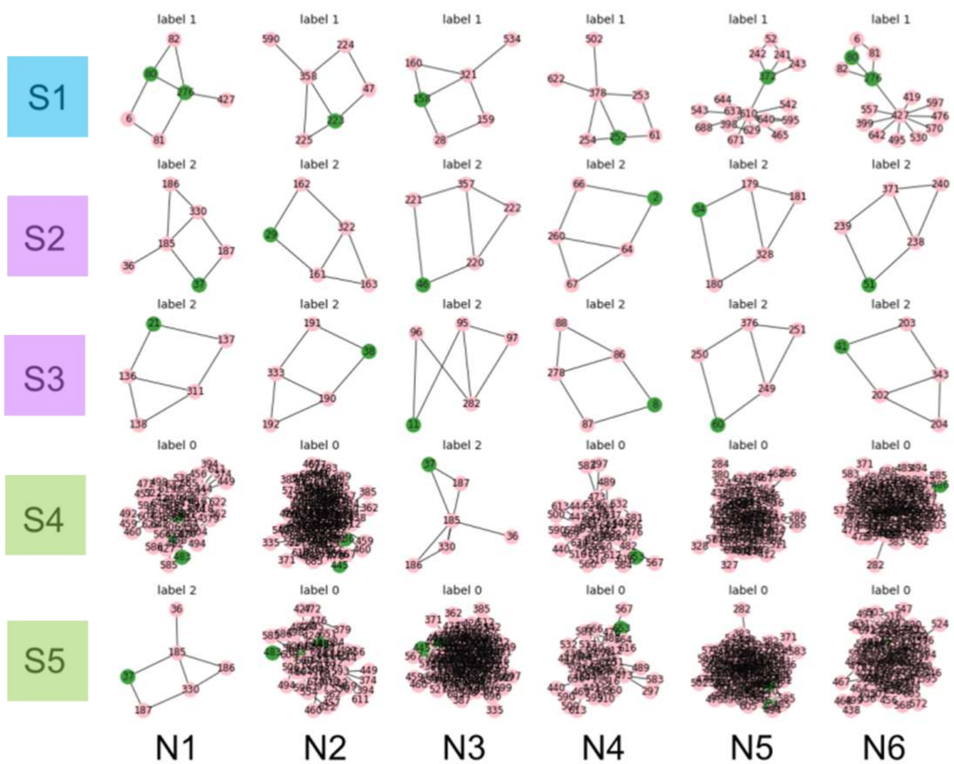


3 motifs – base, circular, flow

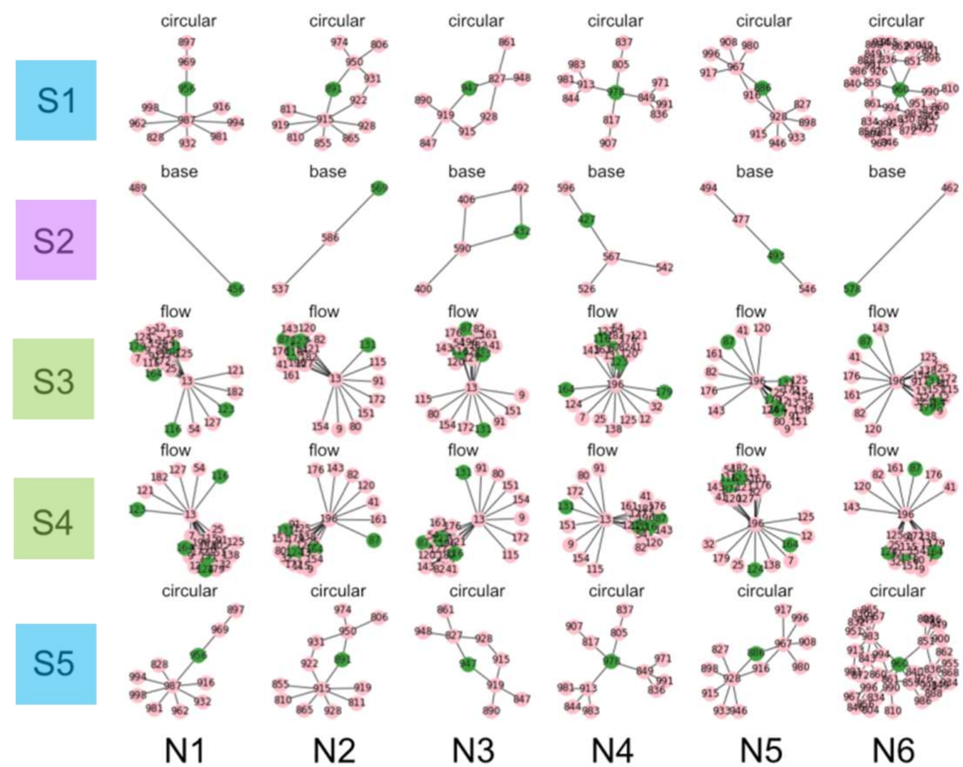
Interpretability

Selected Results

Static BA-Shapes Network

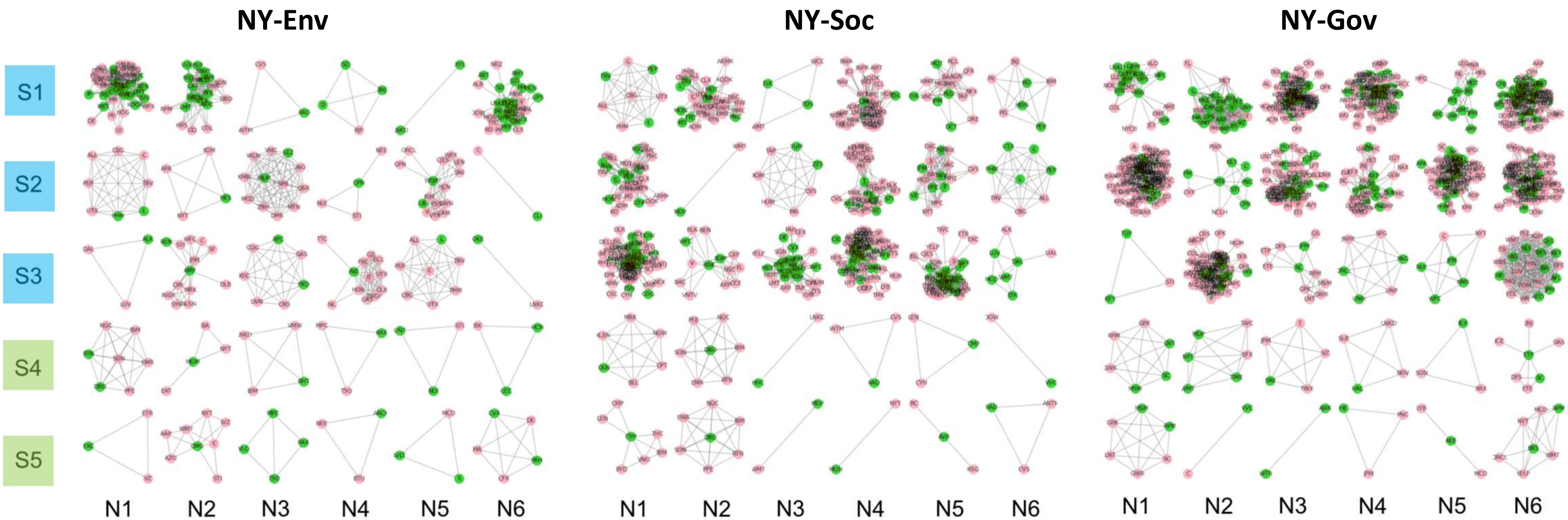


Synthetic Dynamic Network



Interpretability

Selected Results



- S1, S2, S3 focus on nodes in complex and densely inter-connected network patterns, S4 and S5 focus on nodes in relatively simpler network patterns

Conclusion

- Intersection of multiple research areas - learning ***dynamic multimodal networks*** involves ***time-series***, ***multimodal***, and ***network*** *learning*
  - Showed that effective learning of dynamic multimodal networks important for real world tasks
- Challenging due to ***heterogeneity*** and ***dynamic nature*** of information
  - Research across three parts addressed key challenges, proposed ***nine models***
  - Collected ***24 distinct datasets*** for experiments as dynamic multimodal network datasets not common
  - Experiments focused on **important HCI, financial, sustainability domains**, but proposed models can be applied to other tasks and domains (e.g., DynScan on different synthetic datasets)
- ***Interpretability*** of models challenging due to ***diversity of information***, but important to understand model predictions for the same reason
  - MAAN, EMAAN, HAMP, AHAMP, and DynScan models designed for interpretability

Future Research

- Additional dynamic multimodal network **datasets**
  - Many possibilities, e.g., creative, social media datasets
- Learning dynamic multimodal networks for **tasks** in other domains
  - Generative or predictive tasks in creative (3D, video), economic (micro- & macro-economic), social media domains
- **Interpretability** of dynamic multimodal network models
  - E.g., model agnostic methods such as counterfactuals for what-if analysis