MULTIDIMENSIONAL NETWORK ANALYSIS OF CUSTOMER PREFERENCES IN ENGINEERING DESIGN

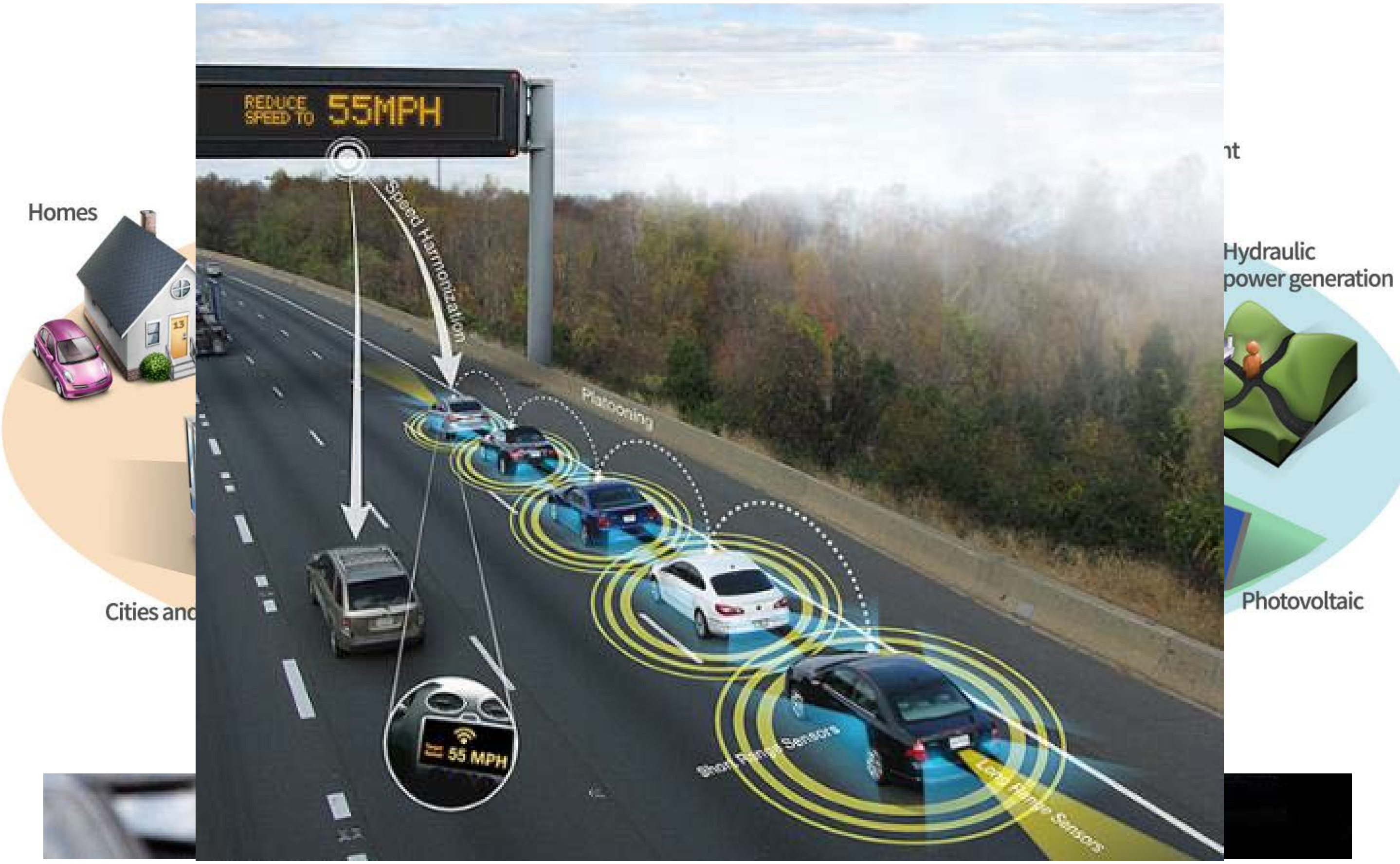
Wei Chen[a] | Mingxian Wang[c] | Zhenghui Sha[a] | Sophia Fu[b]
Yun Huang[b] | Noshir S. Contractor[b] | Yan Fu[c]

Integrated DEsign Automation Laboratory (IDEAL)

a Department of Mechanical Engineering, Northwestern University
b Science of Networks in Communities, Northwestern University
c Global Data Insight & Analytics, Ford Motor Company

Northwestern | McCORMICK SCHOOL OF ENGINEERING

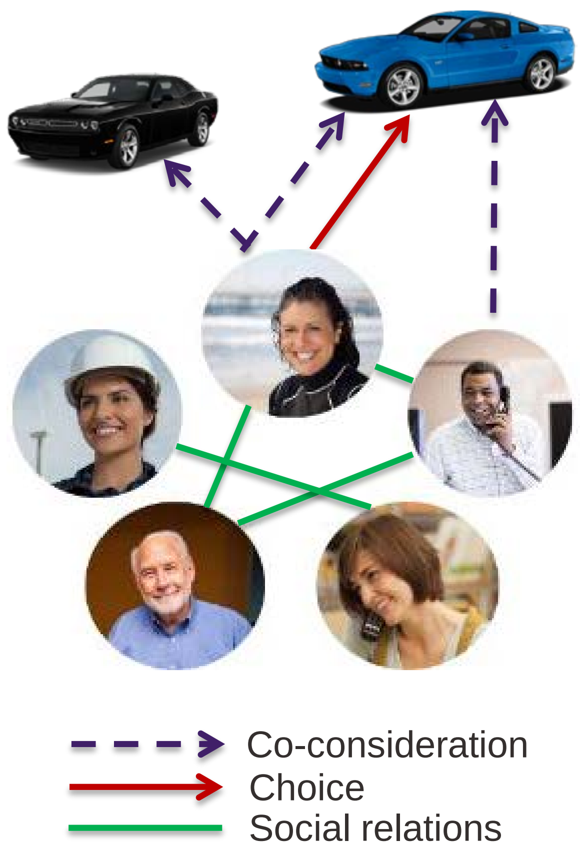
Complex Sociotechnical Systems



Smart Transportation System

(Image source: http://www.its.dot.gov/image_gallery/image36.htm)

Analyzing Customer Preferences



Decision-making behaviors

- The co-consideration behavior
- The choice behavior

Decision-making factors

- Products' attributes, e.g., color, price, etc.
- Demographics, e.g., age, income, etc.
- Usage context; Policy and incentives
- Social influence

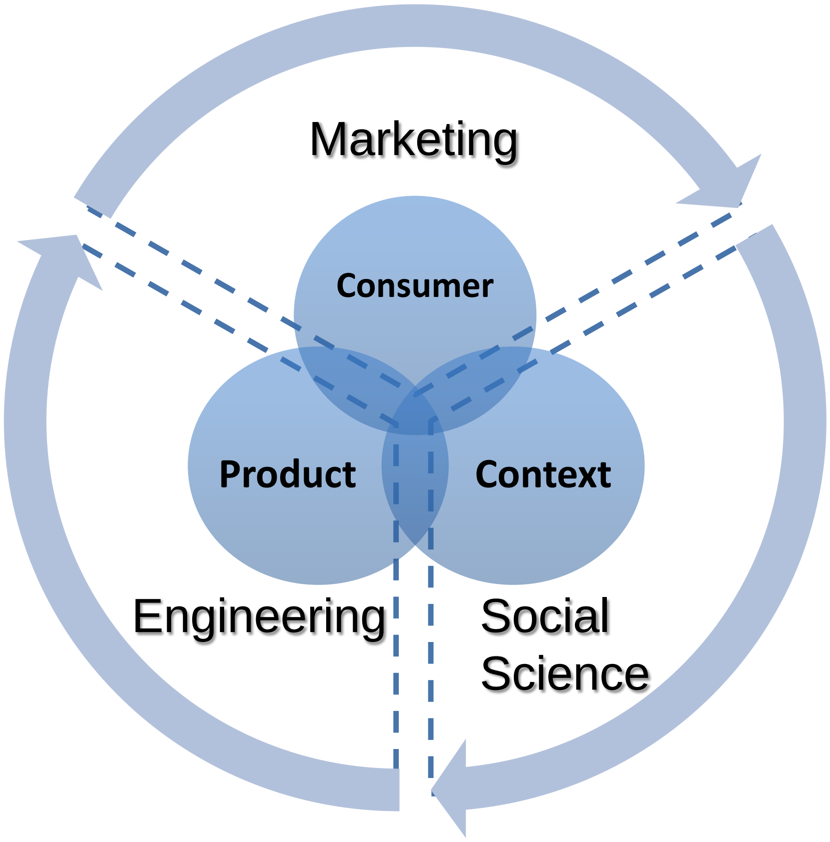
Why it is important?

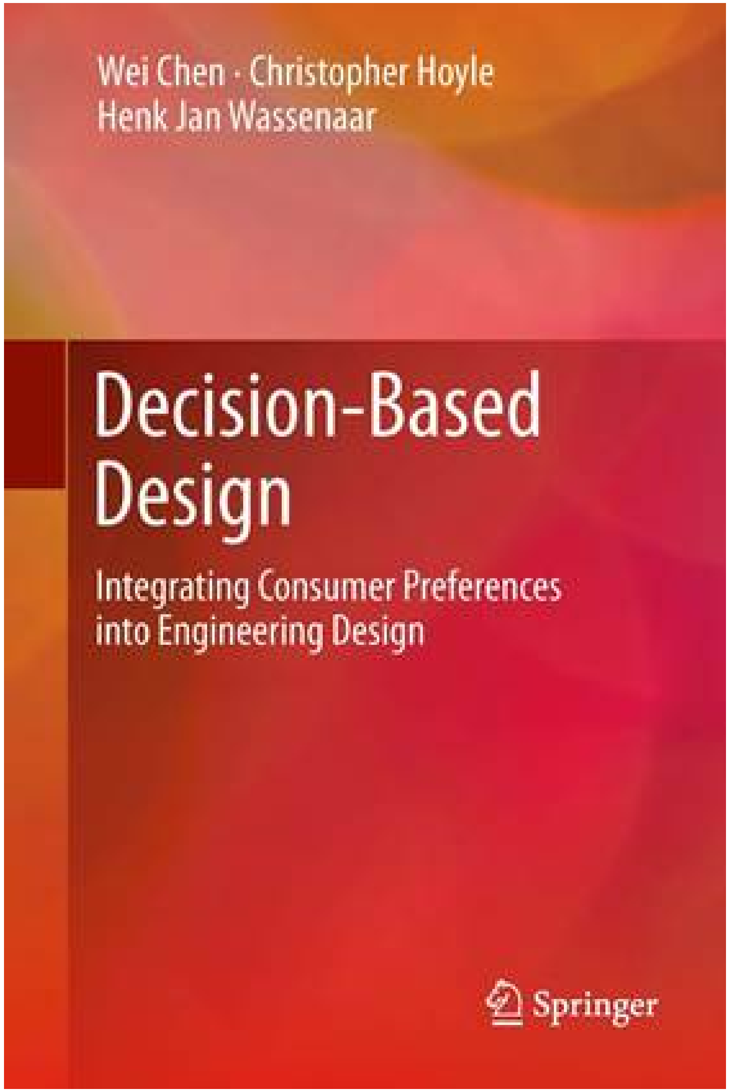
- Support design decisions
- Understand market for strategic planning
- Set right incentives

Modeling Customer Preferences

Decision-based Enterprise-Driven Design

(Chen, Hoyle, Wassenaar, 2013 Springer)





Utility-based Choice Modeling

Discrete Choice Analysis (DCA), rooted in econometrics, used to estimate consumer choice among competing products.

Demographics, Usage S

Product Attr. A

Customer Utility W

Customer Choice

Market Demand

A Customer-desired product attributes
S Demographic & usage attributes

Observed utility

$$W_{in} = \beta_{0i} + \boldsymbol{\beta}_{1i}\mathbf{S}_n + \boldsymbol{\beta}_2\mathbf{A}_i + \boldsymbol{\beta}_3(\mathbf{S}_n \cdot \mathbf{A}_i)$$

$$U_{in} = W_{in} + \varepsilon_{in} \quad \text{if } i \in CS_n$$

Random error

Choice set

$$\Pr_n(i : CS_n) = \frac{e^{W_{in}}}{\sum_{j \in CS_n} e^{W_{jn}}}$$

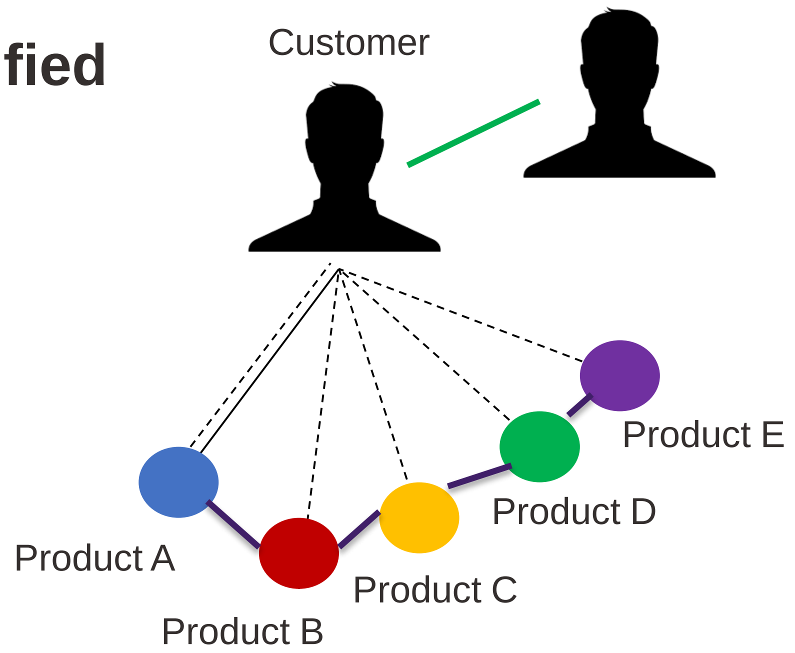
Choice Prob. in Multinomial Logit

$$Q(i) = \sum_{n=1}^{N} \Pr_n(i)$$

Estimated demand of product *i* for market with population N

Limitations of Discrete Choice Analysis (DAC)

- Choice set needs to be prespecified
- Independence of Irrelevant Alternative (IIA) assumption
- Rationality assumption (independent decision maker)
- Vulnerable to attributes collinearity



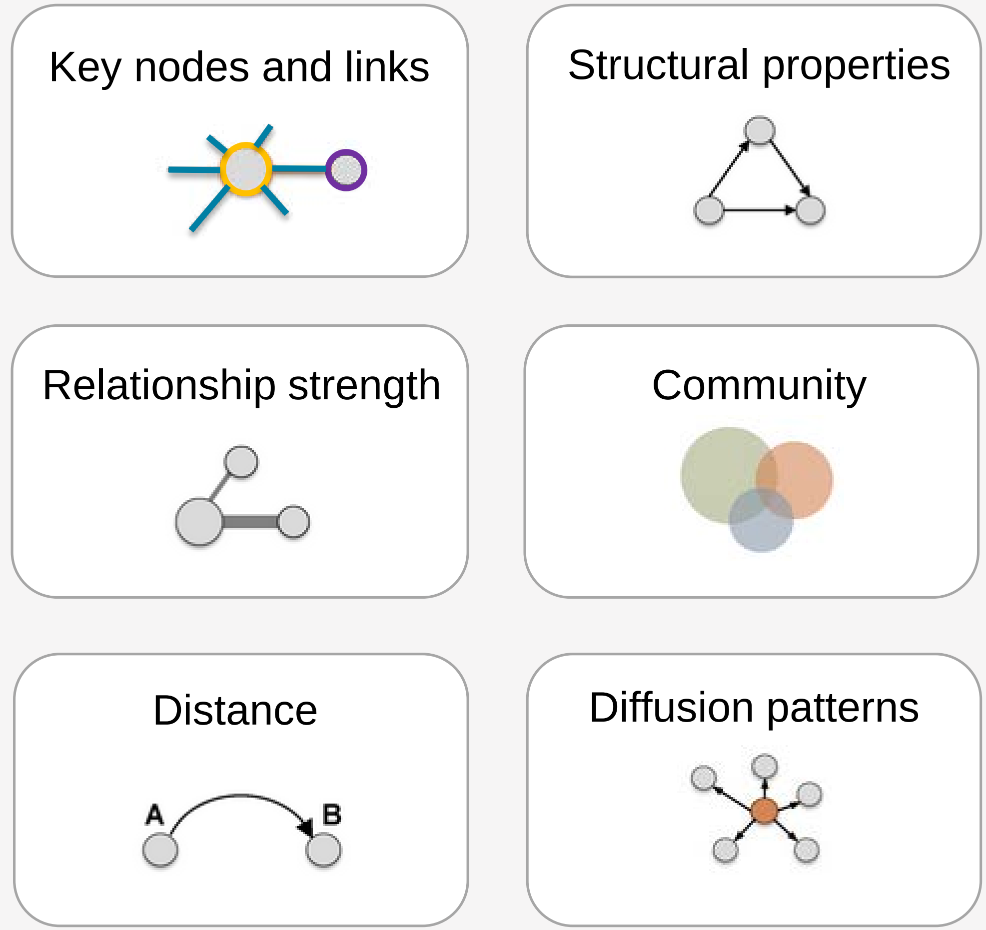
Fundamentals of Network Analysis

Network Structure Analysis

Nodes: individual entities, e.g., customers, vehicles, etc.

Links: complex relations, e.g., social interaction, choice behavior, co-consideration, etc.

Graph: the system structure, e.g., the customer-product systems



Effective in modeling the *interconnectivity* and *interdependency* among individual entities.

Advance of Network Models

Degree-based models

- Erdos-Renyi model
- Small world model
- Barabasi-Albert model

Agent-based models

- Space embeddedness
- Role playing games

Stochastic network models

- Exponential Random Graph

Network realization

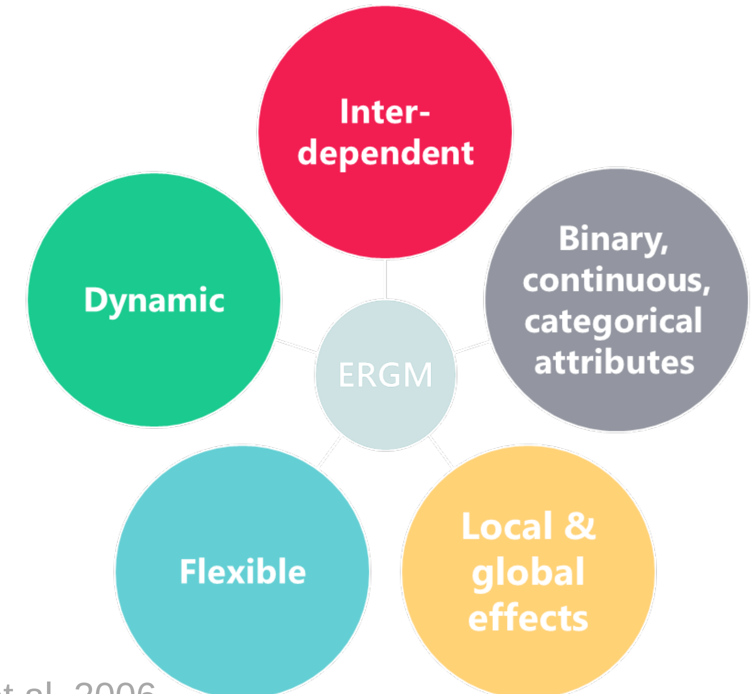
Coefficients

Network Statistics

$$P_{\boldsymbol{\theta}}(\mathbf{Y}=\mathbf{y})=\frac{1}{c(\boldsymbol{\theta})}\exp\{\boldsymbol{\theta}^{\mathbf{T}}\mathbf{z}(\mathbf{y})\}$$

Observed Network

Normalizing Factor

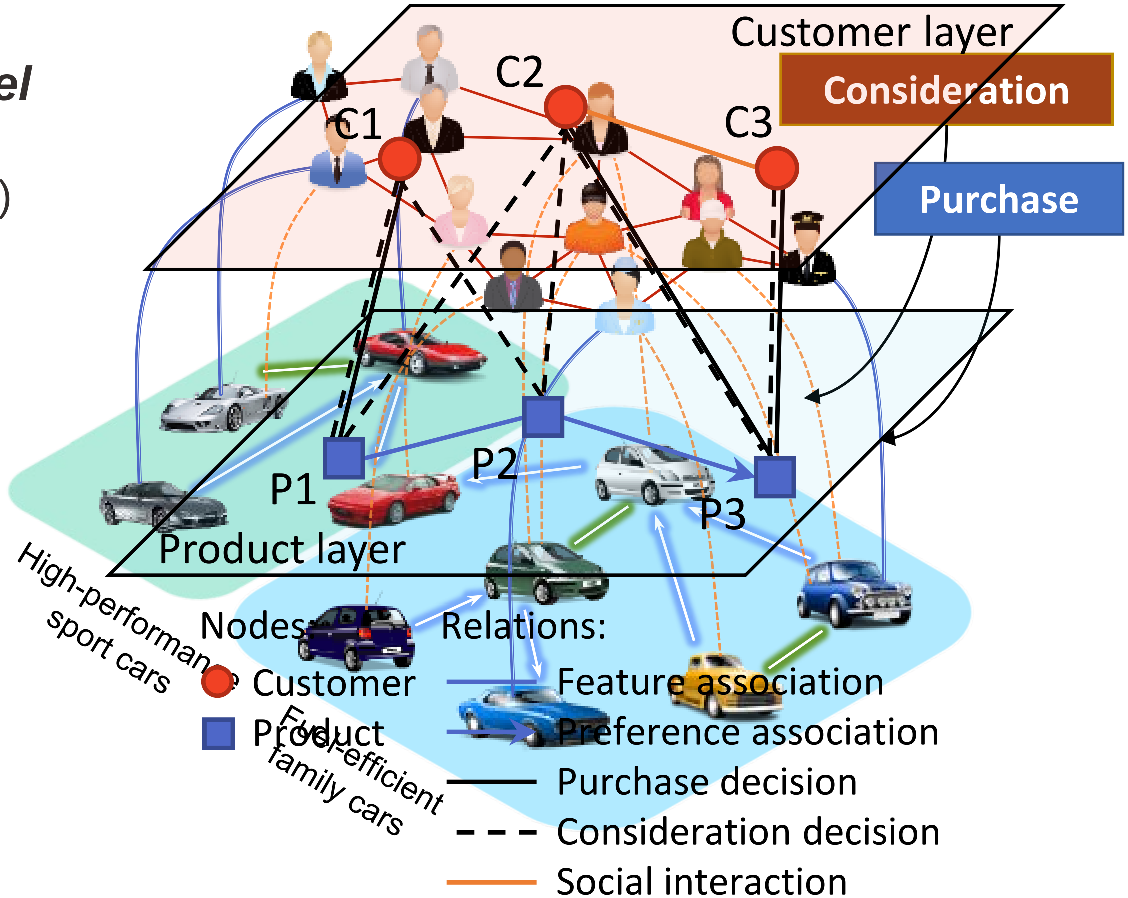


Monge and Contractor 2003
Contractor, Monge, and Leonardi 2011
Wang et al. 2013, Lusher et al. 2012, Robins et al. 2007, Snijders et al. 2006

Multidimensional Customer-Product Network

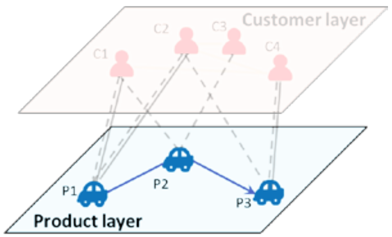
***Multitheoretical multilevel (MTML)* framework**
([Monge and Contractor, 2003](#))

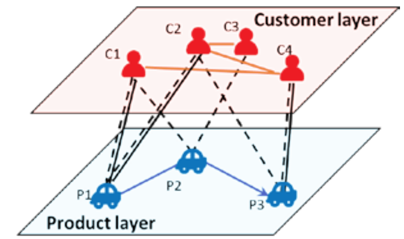
- Self-interest
- collective action
- social exchange
- balance
- homophily
- proximity
- contagion
- co-evolution
- etc.



Wang, M., Huang, Y., Contractor, N., Fu, Y., and Chen, W., "Modeling Customer Preferences using Multidimensional Network Analysis in Engineering Design", *Design Science*, 2016.

Research Topics and Methods





Product Association Network

- Market segmentation and competition
- Key attributes drivers
- Cross-shopping decisions

Descriptive Analysis

e.g., community detection, degree analysis, Joint correspondence analysis (JCA), etc.

Model-based Analysis

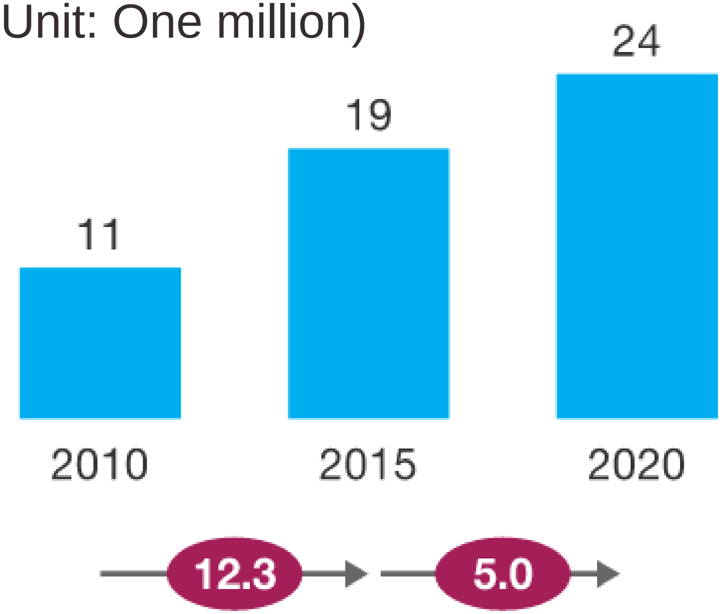
e.g., Multiple Regression Quadratic Assignment Procedures (MRQAP), Exponential Random Graph Model (ERGM)

Multidimensional Customer-Product Network Analysis

- Attribute effects
- Network structure effects
- social network effect

Project Context

Annual Passenger Car Sales in China
(Unit: One million)



Compound annual growth rate in percentage

- Regional differences
- Diverse preferences
- Intense competitions
- Social Influences

Market facts:

- China surpassed US to become the No.1 auto market in 2010.
- China is expected to exceed North America and Europe to become the No.1 area market in 2020.

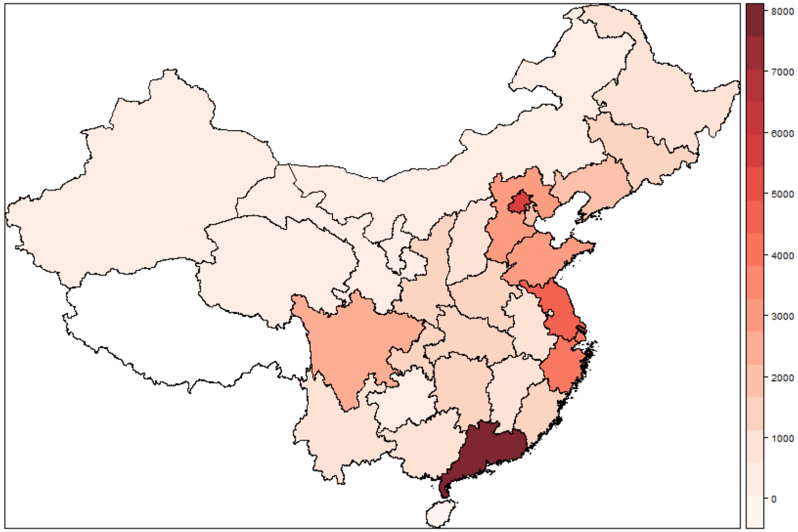
Data Source

New Car Buyers Survey (NCBS) 2013

- ~ 50,000 respondents
- 389 vehicles
- 872 variables

Covered factors

- Purchased vehicle
- Considered alternative vehicles
- Previous owned vehicle
- Vehicle attributes (e.g., body type, engine power)
- Demographics (e.g., age, income)
- Use patterns (e.g., average km per day)
- Perceived vehicle characteristics (e.g., youthful, reliable)

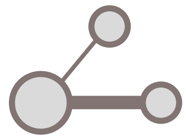


2013 Mainland China, 54 cities in 30 provinces

Product Association Network

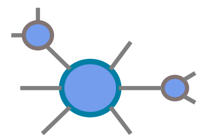
NETWORK LINK

- Undirected, link strength
- Co-consideration



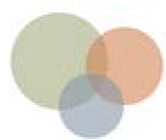
CENTRALITY (SIZE)

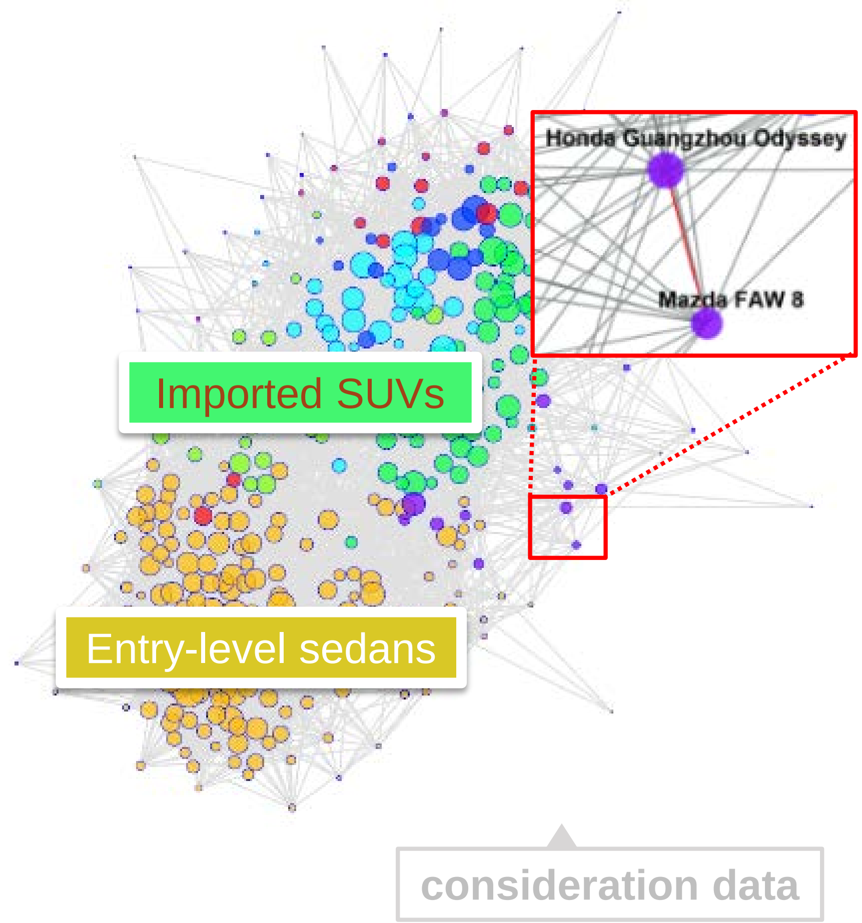
- Identify hubs
- Imply consideration range



COMMUNITIES (COLOR)

- Detect group of products with strong connections
- Imply market segment and aggregated consideration set

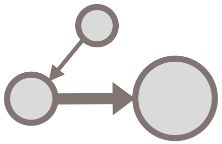




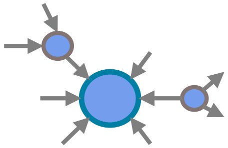
Vehicle Hierarchical Network

NETWORK LINK

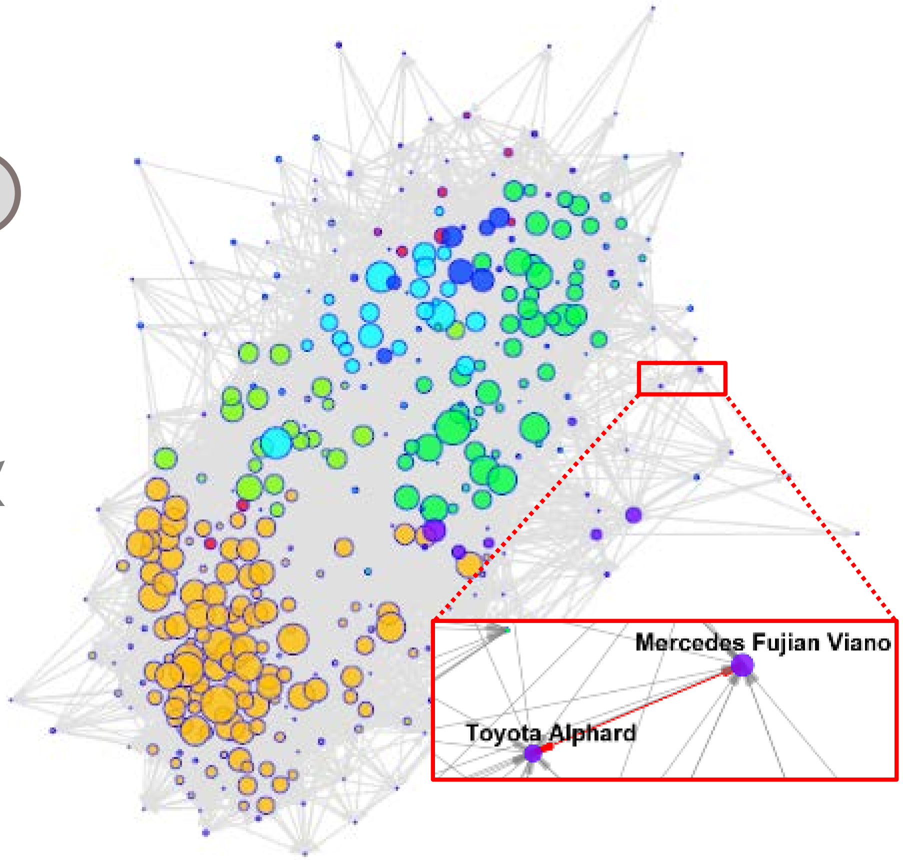
- Directed, valued
- Co-consideration
- Purchase preference



NODE HIERARCHY (SIZE)



- Identify winner products in pair-wise evaluations
- Imply product competitiveness under co-consideration



consideration & purchase data

Unidimensional Vehicle Network based on Costumers' Co-Considerations

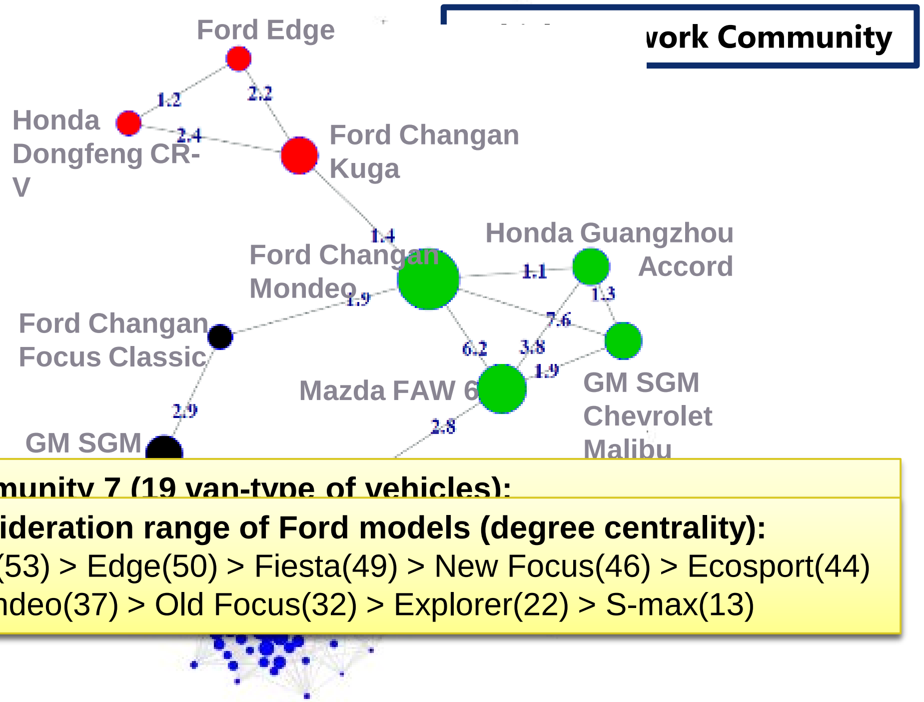
Evaluation Metric

$$lift = \frac{P(i \cap j)}{P(i) \cdot p(j)}$$

Link Generation

$$Edge(i,j) = \begin{cases} 1, & if\ lift\ (i,j) > cutoff \\ 0, & otherwise \end{cases}$$

- A vehicle network that connects the co-considerations of cars (2013 data)



Community 7 (19 van-type of vehicles):

Consideration range of Ford models (degree centrality):
Kuga(53) > Edge(50) > Fiesta(49) > New Focus(46) > Ecosport(44) > Mondeo(37) > Old Focus(32) > Explorer(22) > S-max(13)

Joint Correspondence Analysis (JCA)

Identify key attributes drivers to the formation of network communities

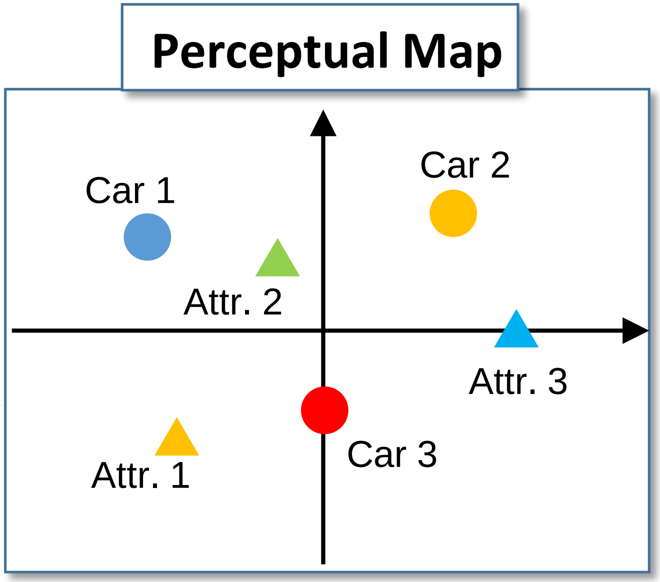
	Car 1	Car 2	Car 3	….	Attr. 1	Attr. 2	
Person 1	0	0	1	…	0	0	…
Person 2	1	0	0	…	0	1	…
Person 3	1	0	0	…	1	0	…
Person 4	0	1	0	…	0	1	…
….	…	…	…	…	…	…	…

Dimension reduction by matrix transformation

Calculation of principal coordinates and inertia

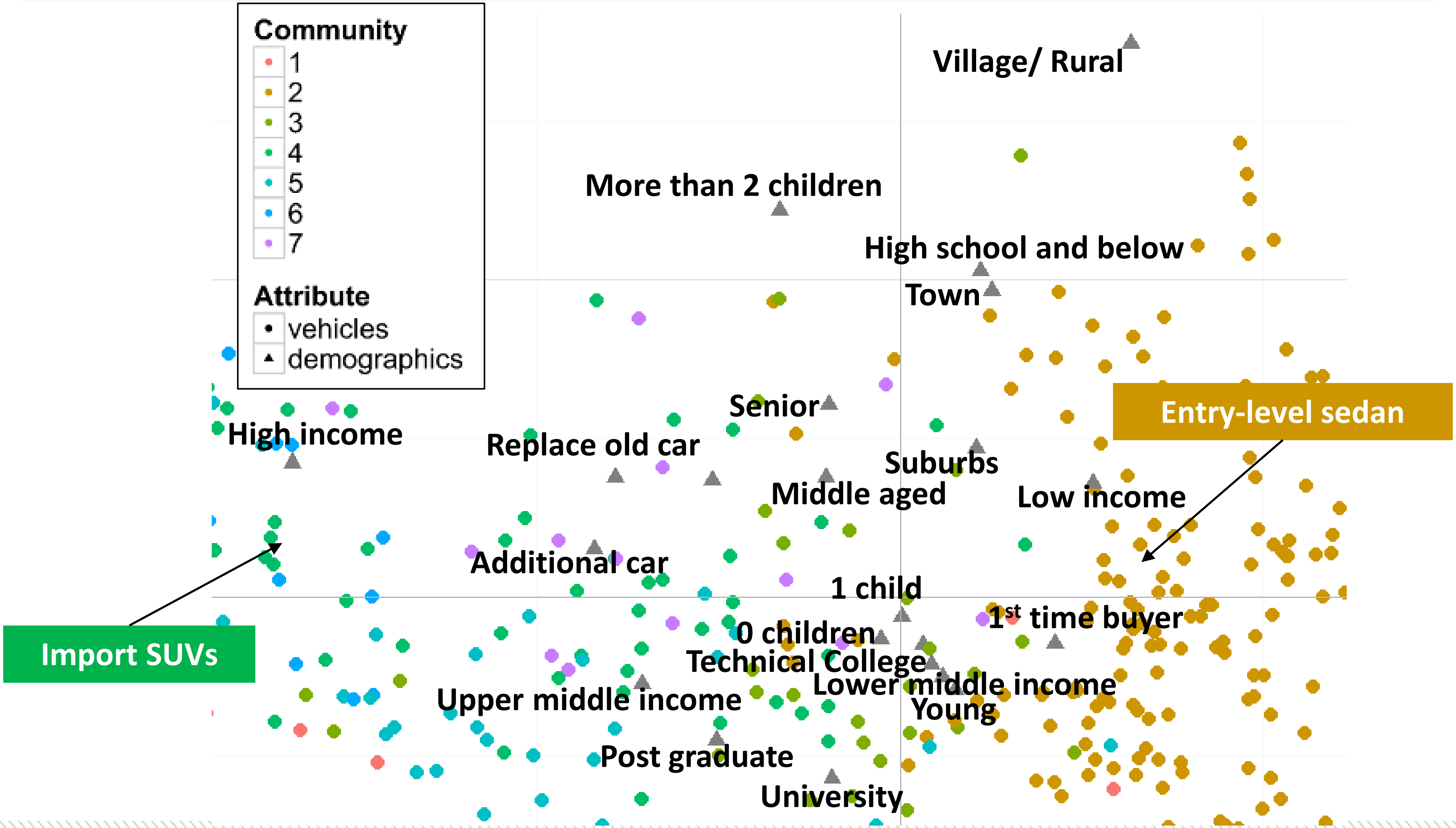
Plot first 2 principal coordinates

- A multivariate approach
- Generate a visual perceptual map
- Development in Literature
  - CA → MCA → JCA



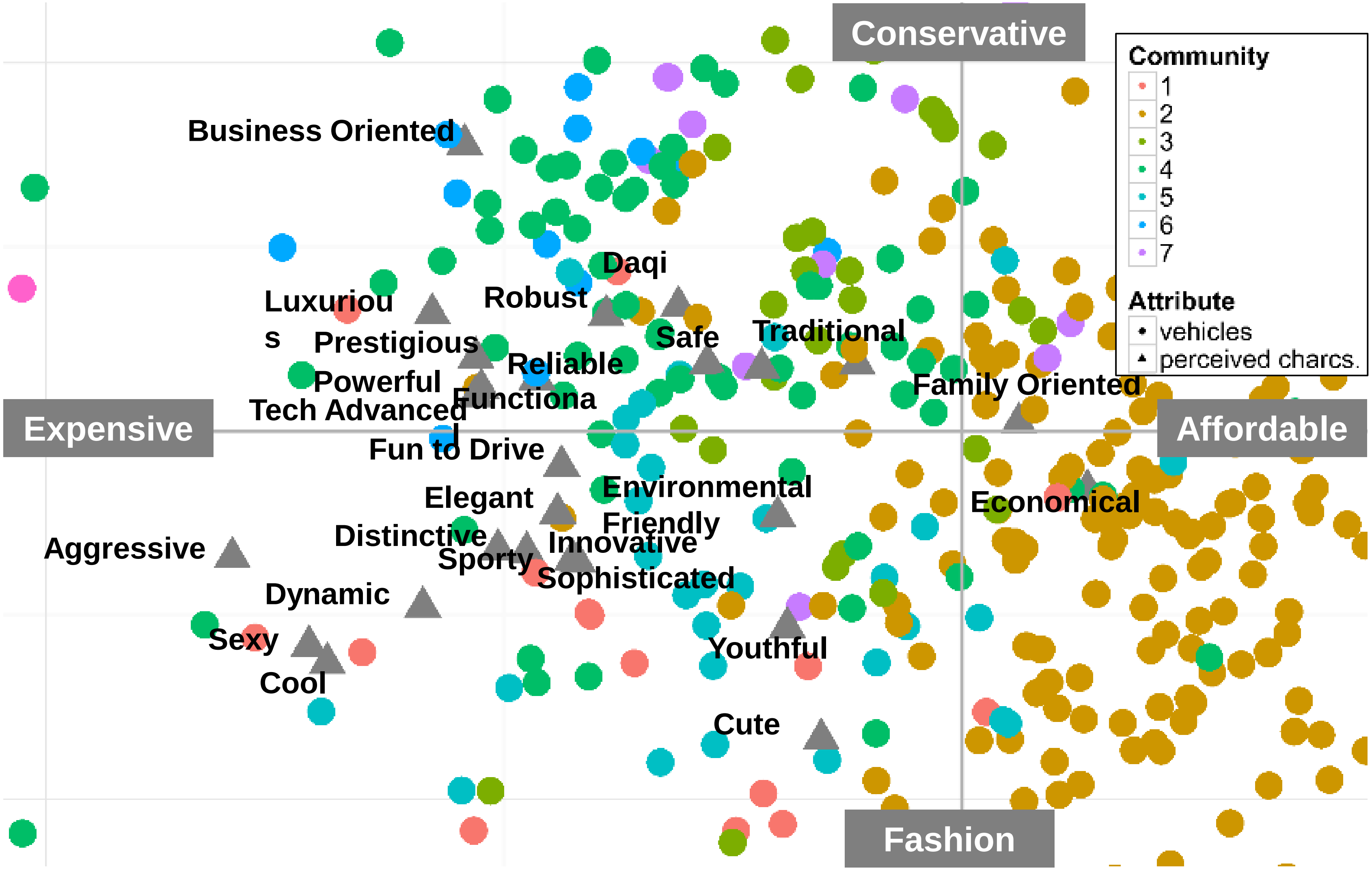
JCA of Vehicles and Demographics

WHAT CUSTOMER DEMOGRAPHICS EXPLAIN PRODUCT COMMUNITIES?



IDEAL

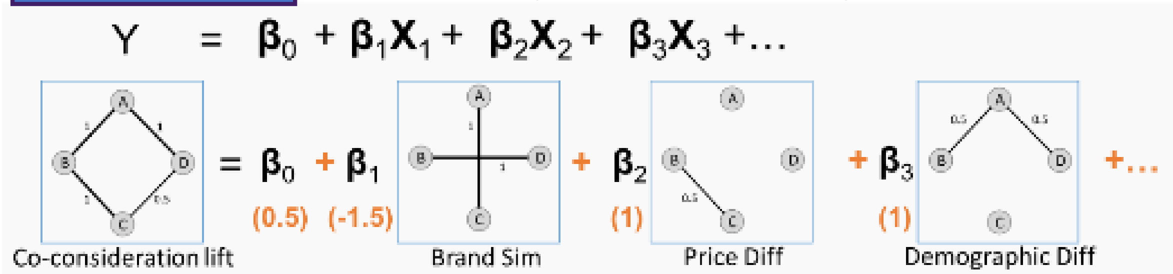
JCA of Vehicles and Perceived Vehicle Char.

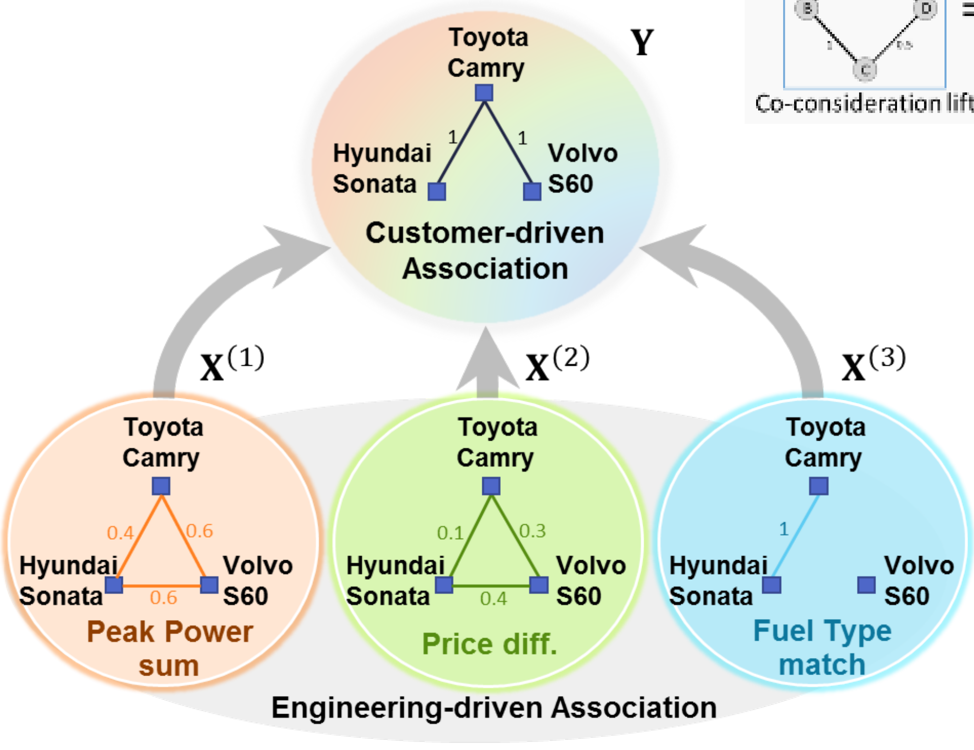


IDEAL

Multiple Regression Quadratic Assignment Procedures

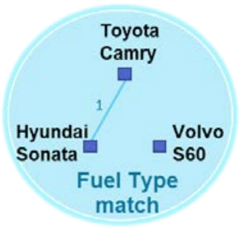
MRQAP: $\text{logit}\left(P_{\boldsymbol{\theta}}\{Y_{ij} = 1\}\right) = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_{ij}$

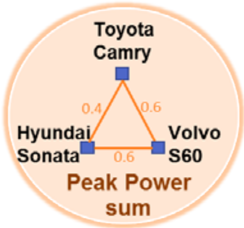




- Predict co-consideration relations using networks formed by attributes
- Explain the impact of similarity and differences on co-considerations
- Relate customer-driven association to engineering-driven associations

Model Configuration and Network Effect

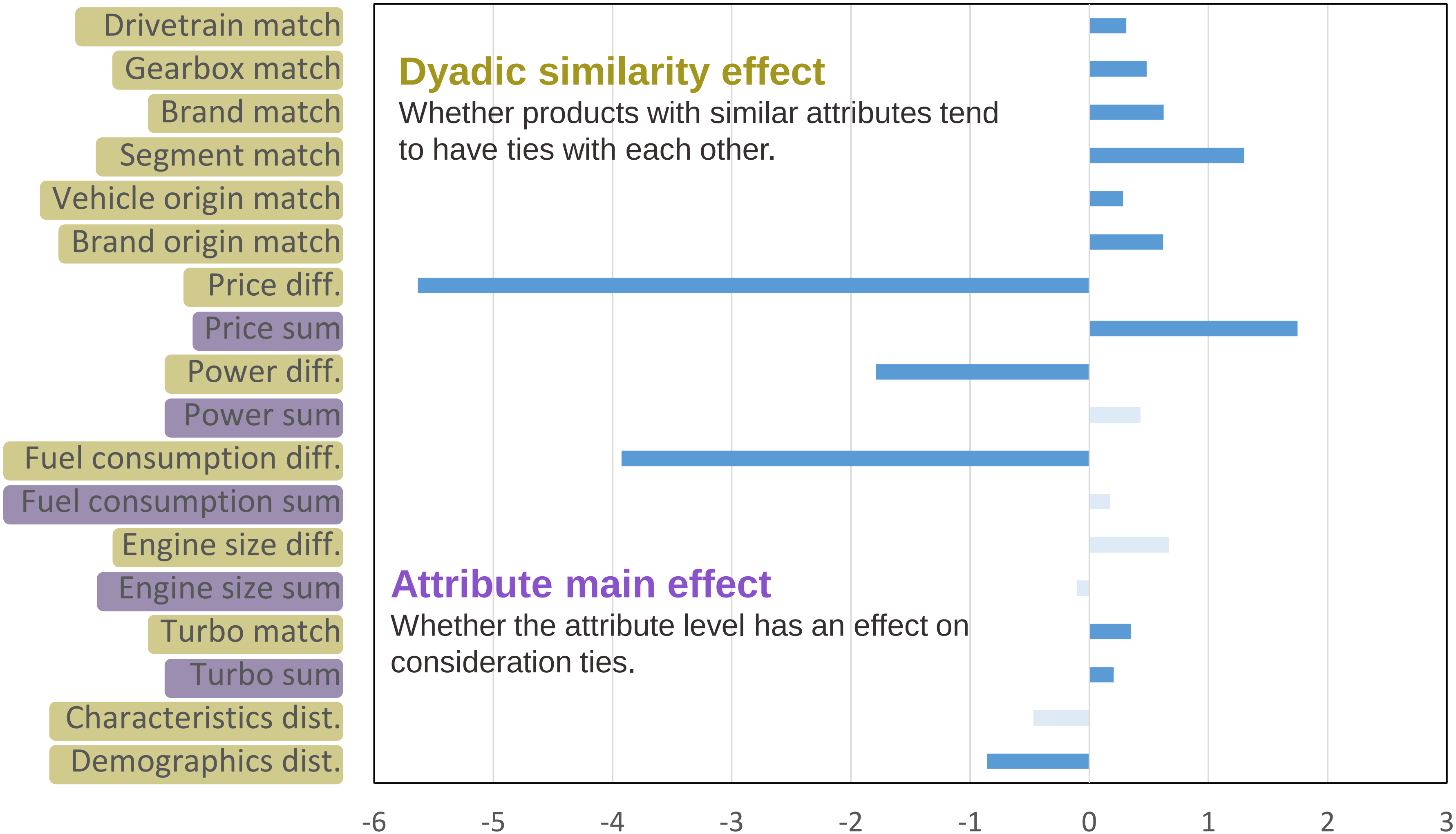




Configuration	Statistic	Network effect
Binary product attributes		
Sum network	$X_{ij} = x_i + x_j$	Attribute-based main effect
Match network	$X_{ij} = I\{x_i = x_j\}$	Homophily effect
Categorical product attributes		
Match network	$X_{ij} = I\{x_i = x_j\}$	Homophily effect
Continuous product attributes (standardized)		
Sum network	$X_{ij} = x_i + x_j$	Attribute-based main effect
Difference network	$X_{ij} = \|x_i - x_j\|$	Homophily effect
Non- product related attributes		
Distance network	$X_{ij} = \|\|x_i - x_j\|\|_2$	Homophily effect

Wang, M., Huang, Y., Contractor, N., Fu, Y., and Chen, W., "A Network Approach for Understanding and Analyzing Product Co-Consideration Relations in Engineering Design", *International Design Conference – Design 2016*, Dubrovnik, Croatia.

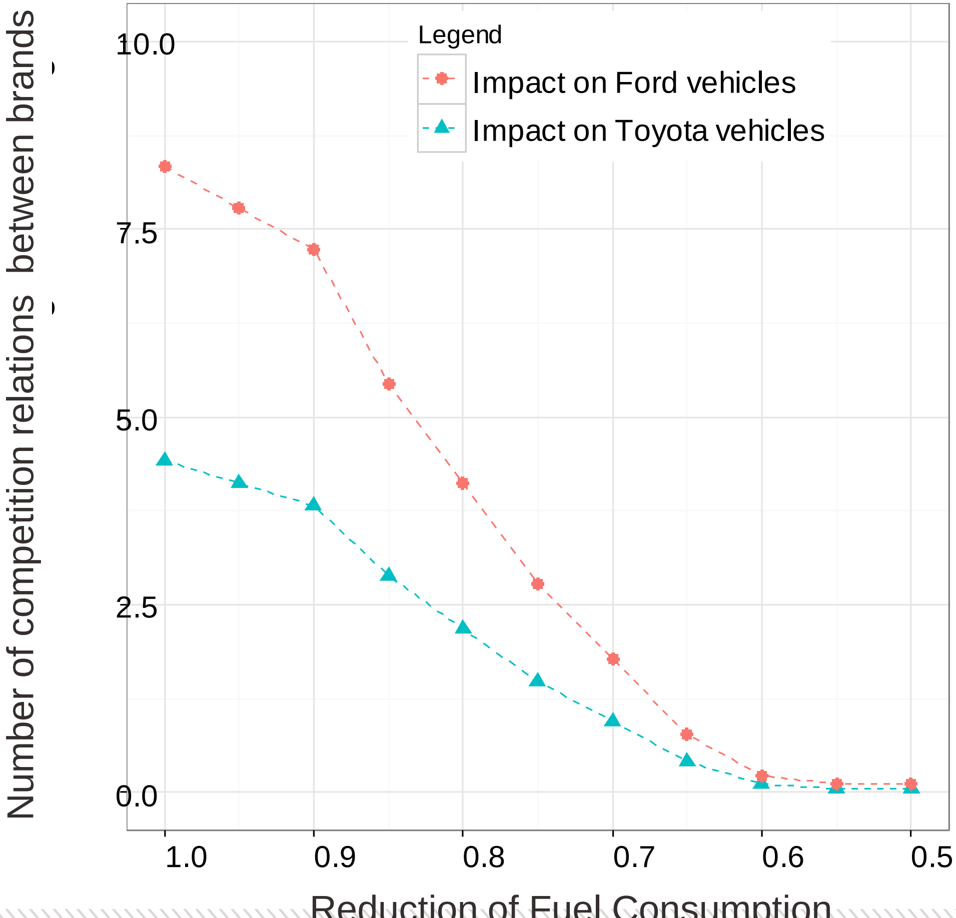
MRQAP Network Modeling Results



Forecast Technological Impacts

Design Scenario: Improve fuel economy

MRQAP: $\text{logit}\left(P_{\boldsymbol{\theta}}\{Y_{ij} = 1\}\right) = \boldsymbol{\beta}^{\mathbf{T}}\mathbf{x}_{ij}$



Wang, M., Sha, Z., Huang, Y., Contractor, N., Fu, Y., and Chen, W., "Forecasting Technological Impacts on Customers' Co-Consideration Behaviors: A Data-Driven Network Analysis Approach" , IDETC2016-60015, *Proceedings of the ASME 2016 International Design Engineering Technical Conferences & Design Automation Conference*, August 21-24, Charlotte, NC.

Comparative Study on Network Models

Network Analysis Model (MRQAP)

- Independent assumption
- Predicting each pair of vehicles
- Vehicle attributes
- Computationally less expensive

Product Association Network

I. Network modeling of customer preferences

II. Network prediction of market competitions under technological changes

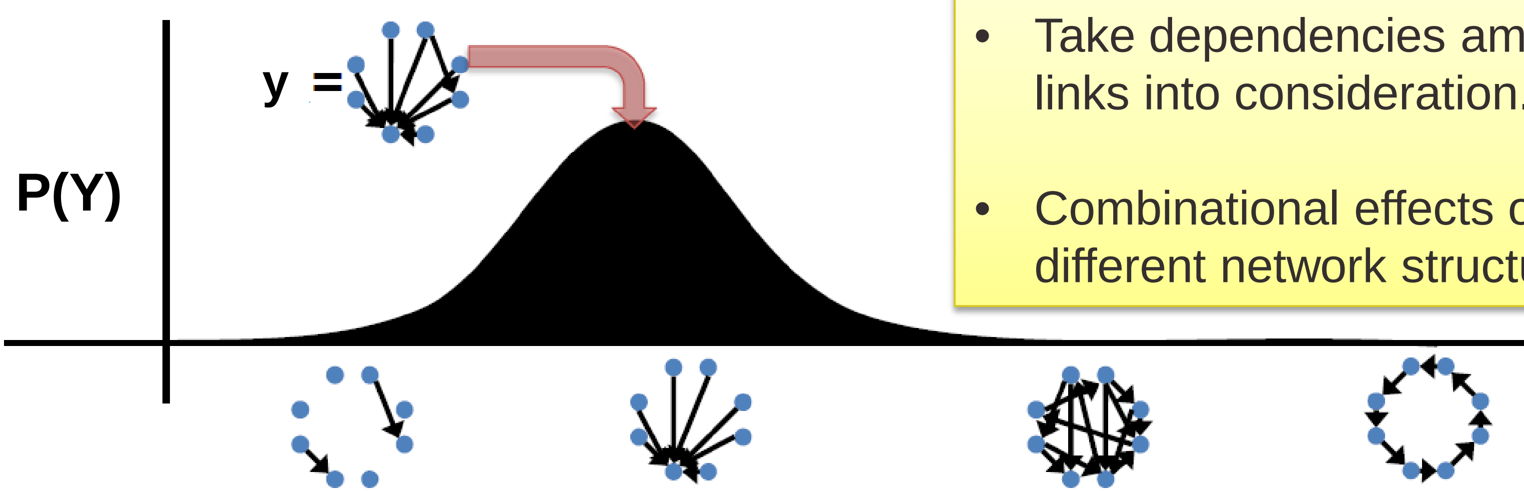
Exponential Random Graph Model (EGRM)

- Dependency among links
- Predicting the whole network
- Network configuration
- Computationally expensive

Exponential Random Graph Model

Exponential Random Graph Model (ERGM)

$$\Pr\left(Y=y\right)=\underbrace{\left(\frac{1}{c}\right)}_{\text{Probability of the graph}}\exp\left\{\underbrace{\sum_{m}\theta_m x_m\left(y\right)}_{\text{Coefficient * Effect}}\right\}$$



Pros:

- Take dependencies among links into consideration.
- Combinational effects of different network structures.

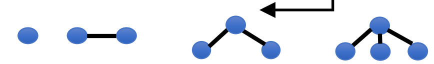
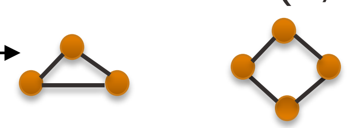
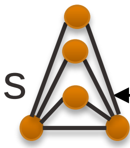
Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007).

Wang, P., Robins, G., Pattison, P., & Lazega, E. (2013).

Modeling Heterogeneity

Structural features of interest !

$$P_\theta(\mathbf{Y}=\mathbf{y}) = \frac{1}{c(\boldsymbol{\theta})}\exp\{\boldsymbol{\theta}^T \mathbf{x}(\mathbf{y})\}$$

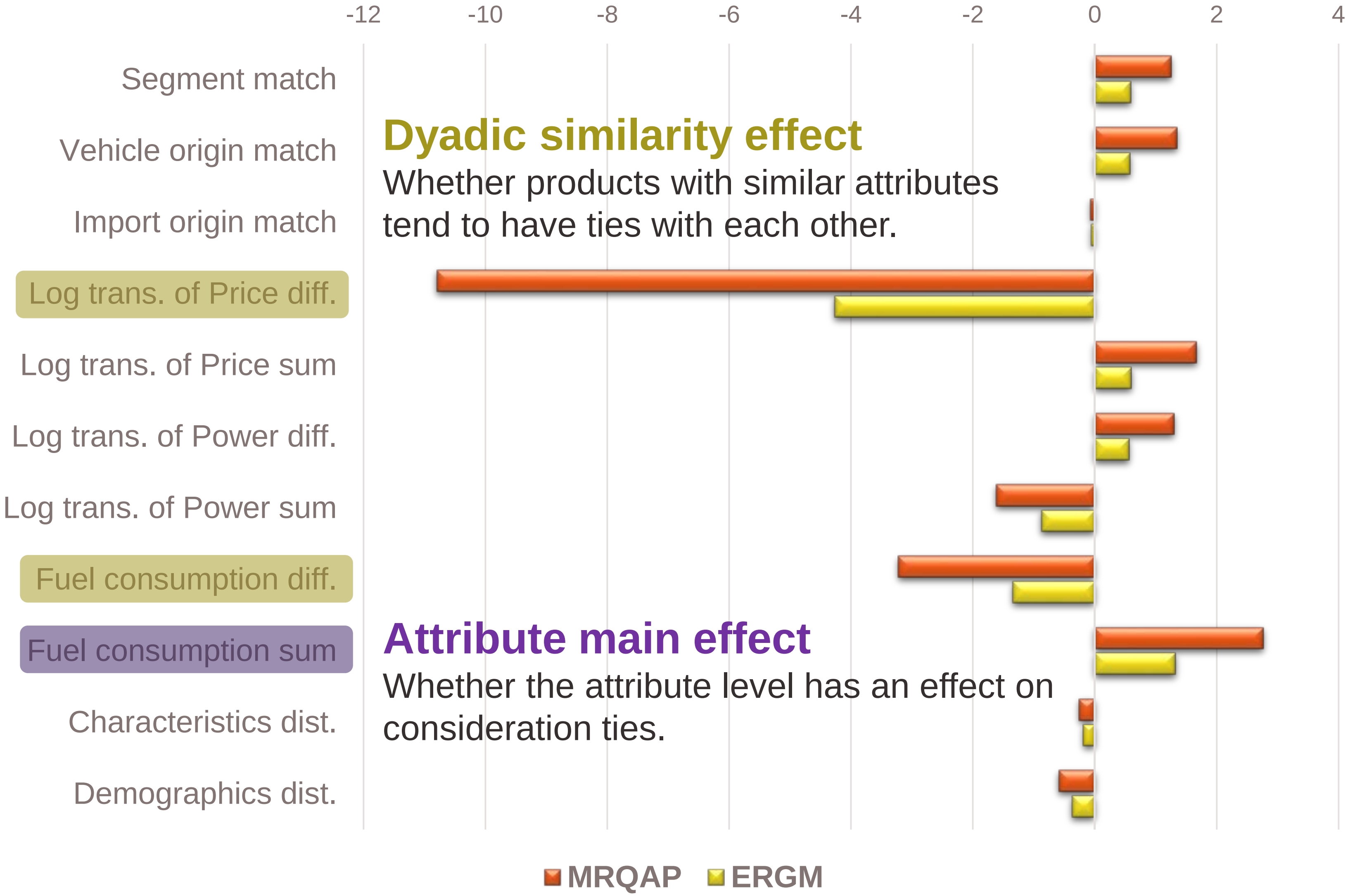
Attributes of nodes	Heterogeneity in – Customer income, age, etc. – Product price, performance, etc.
Attributes of links	Heterogeneity in – Types of relations –Time duration
Network Configurations	Degree distributions (or stars)  Cycle distributions (2, 3, 4, etc.)  Shared partner distributions 

Whether a car is co-considered with other two cars

Whether three cars are co-considered with each other (three-way competition)

Whether two cars are co-considered with many other two cars

Model Results



Dyadic similarity effect

Whether products with similar attributes tend to have ties with each other.

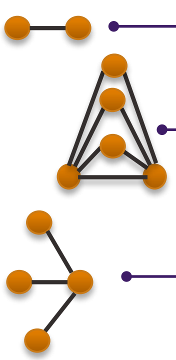
Attribute main effect

Whether the attribute level has an effect on consideration ties.

Insights from ERGM Model

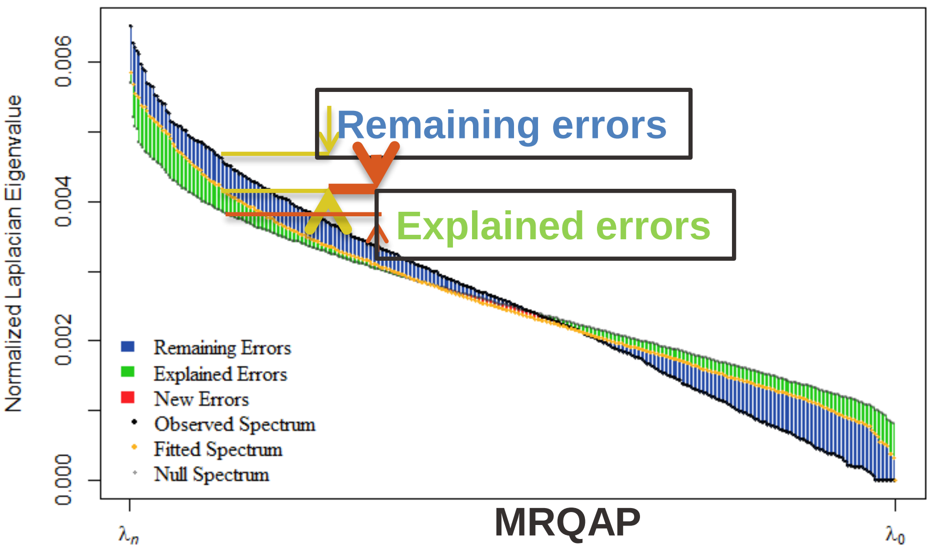
Vehicle Attribute Network

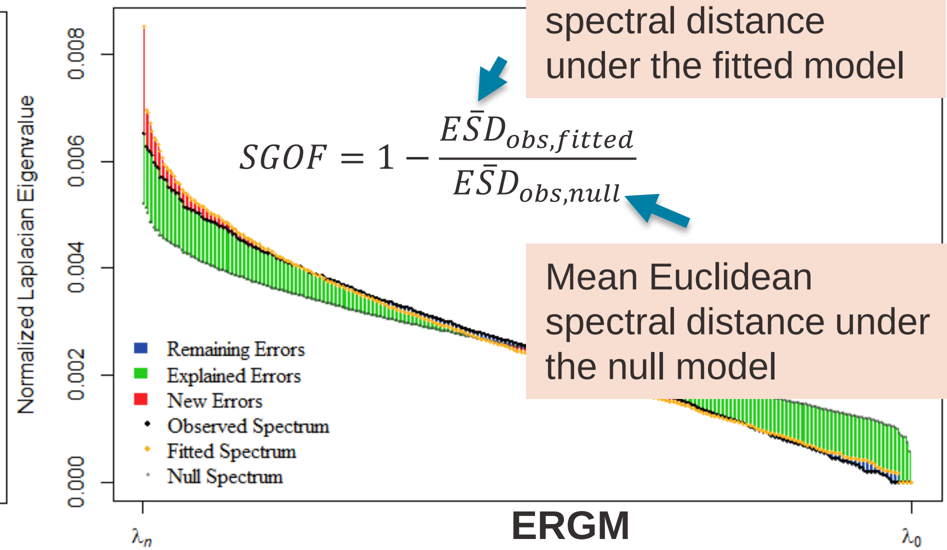
Input attr. name	Input attr. type	Coeff. β
Segment match	Categorical	0.590*
Vehicle origin match	Categorical	0.578*
Import origin match	Categorical	-0.0618
Log trans. of Price diff.	Numerical	-4.269*
Log trans. of Price sum	Numerical	0.601*
Log trans. of Power diff.	Numerical	-0.876*
Log trans. of Power sum	Numerical	0.563*
Fuel consumption diff.	Numerical	-1.344*
Fuel consumption sum	Numerical	1.324 *
Characteristics dist.	Perceived char.	-0.188
Demographics dist.	Customer attr.	-0.374*
edges	Network statistics	-7.790*
Shared partner distribution	Network statistics	0.681*
Degree distribution	Network statistics	2.317*
Overall model fit (Null deviance: 104618)		BIC: 13997



Models Evaluation – Spectral Goodness of Fit (SGOF)

1000 simulations	MRQAP	ERGM
Spectral Goodness of Fit *[Jesse Shore, Benjamin Lubin, 2015]*	**0.35** (5th, 95th quantiles: 0.28, 0.42)	0.69 (5th, 95th quintiles: 0.60, 0.76) 0.67 (5th, 95th quintiles: 0.54, 0.78) (2M burn-in)





Mean Euclidean spectral distance under the fitted model

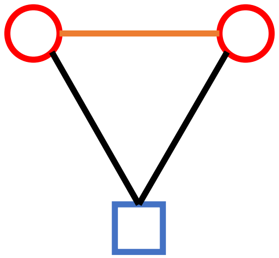
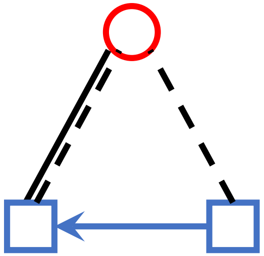
$$SGOF = 1 - \frac{E\bar{S}D_{obs,fitted}}{E\bar{S}D_{obs,null}}$$

Mean Euclidean spectral distance under the null model

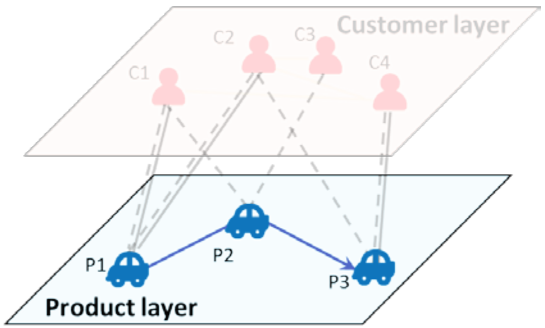
The SGOF measures the amount of observed structure explained by a fitted model.

Modeling Social Influence

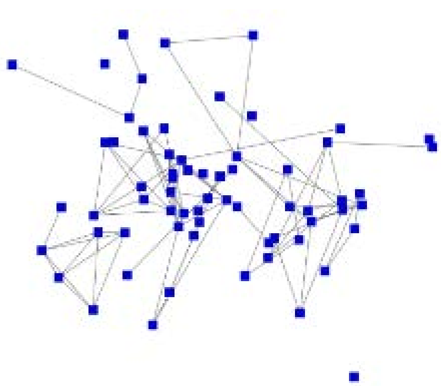
$$P_\theta(\mathbf{Y}=\mathbf{y}) = \frac{1}{c(\boldsymbol{\theta})}\exp\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x}(\mathbf{y})\}$$

Peer effect		Customers tend to choose the product that their "peers" recommended, either through use or discussion.
Crowd effect		When comparing two products under consideration, a customer is more likely to choose the one favored by majority of customers.

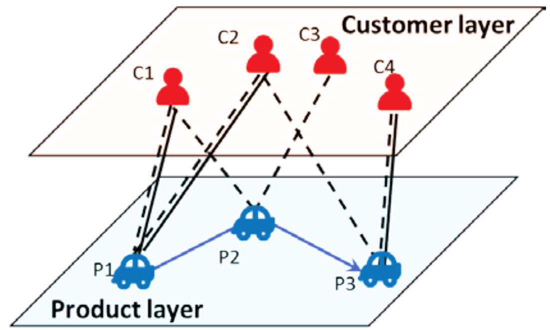
Luxury Vehicle Preferences in Central China



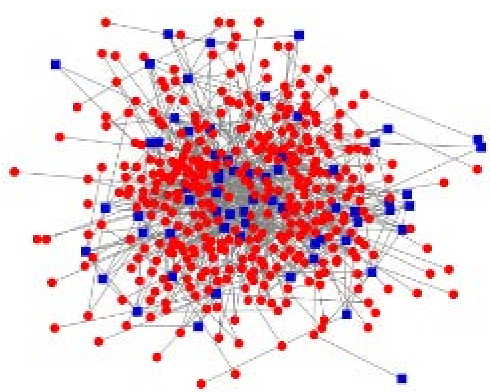
Product feature similarity associations



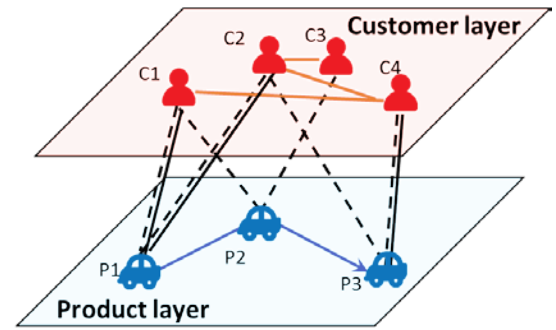
Created by vehicle features and attributes



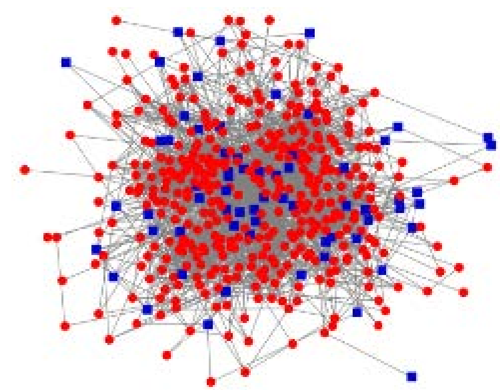
Consideration preference decision added



Defined by consideration decisions in survey data



Customer social interactions added



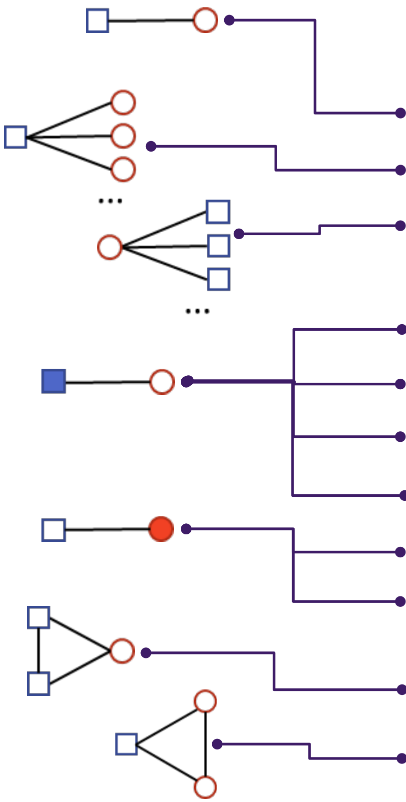
Simulated using the small-world network model

ERGM Modeling Results

Only attribute effects

Structural effects added

Social influence effect added



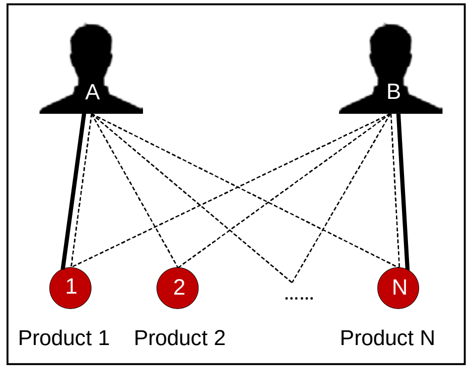
Effects	Model 1	Model 2	Model 3
Pure structural effects			
Density	-7.0314*	-9.1009*	-8.9648*
Product popularity		6.4955*	6.5123*
Consideration range		-1.4036*	-1.3199*
Attribute-relation main effect			
Price	-0.0346*	-0.0194	-0.0182
Turbocharger dummy	1.2796*	1.0617*	0.9056*
Engine capacity	0.2809*	0.2356*	0.1871
Fuel consumption	0.1581*	0.1270*	0.1162*
First-time buyer dummy	-0.2343*	-0.9745*	-0.9744*
Income	0.0027	0.0102	0.0125*
Cross-level effects			
Customer consider similar products		0.9930*	0.9704*
Peer influence in consideration			0.4524*
Model Fit			
AIC	5148	4851	4795
BIC	5205	4932	4884

Research Contributions

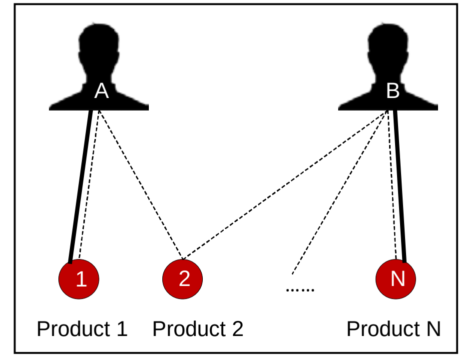
- Employed descriptive network analysis to provide visualization of structure features of vehicle co-consideration relations and identified key vehicle attributes drivers.
- Employed network models to study the impact of similarity and differences of product features on vehicle co-considerations.
- Illustrated the use of network models to predict the impact of different technologies on vehicle competition.
- Compared different networking modeling techniques
- Established a multidimensional network framework for modeling consumer consideration by taking account both product association and social influence.

On-Going:
Two Stage Consideration-then-Choice Models

Previous Approaches

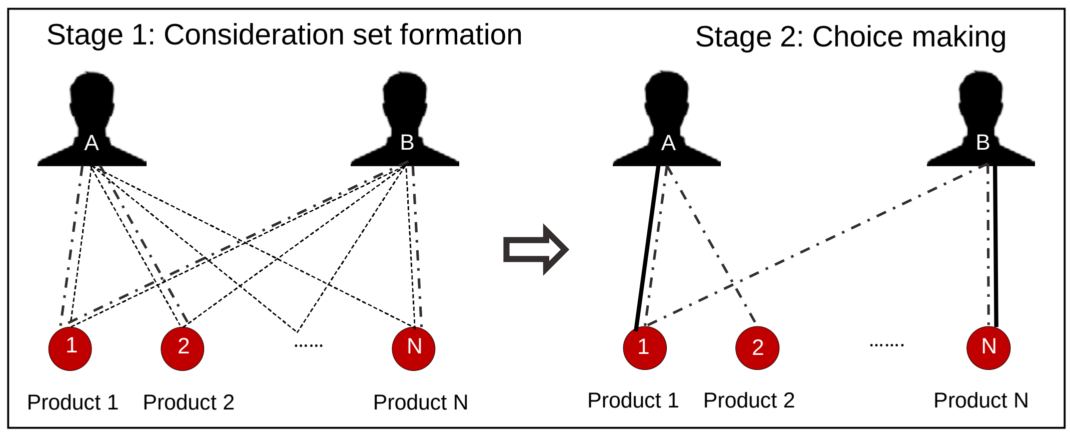


Scenario 1: one-stage choice model assuming customers make decisions among all possible products.



Scenario 2: two-stage choice model assuming each customer makes decisions from a subset of products which is unknown to researchers.

Proposed Approach



Scenario 3: two-stage choice model assuming each customer considers a subset of products first and makes final decisions from it. Researchers have access to both consideration set and the final choices data.

Possible alternative

Consideration

Product choice

Applying Bi-partite ERG Modeling

	Stage 2- Purchase | consideration	Stage 1- Consideration
Edges	14.90**	-10.53**
Market distribution	-3.83**	-4.11**
Price	-0.36	0.30**
Fuel consumption	0.40**	-0.16**
Make origin (US)	-0.84**	0.88**
Make origin (Europe)	-0.33	0.83**
Make origin (Japan)	-0.29	0.00
Make origin (Korea)	-0.50*	0.14
External styling	-1.27**	0.23**
Turbo	0.67**	-1.19**
All wheel drive (AWD)	-1.66**	-0.32**
Auto transmission	-0.48	-0.53**

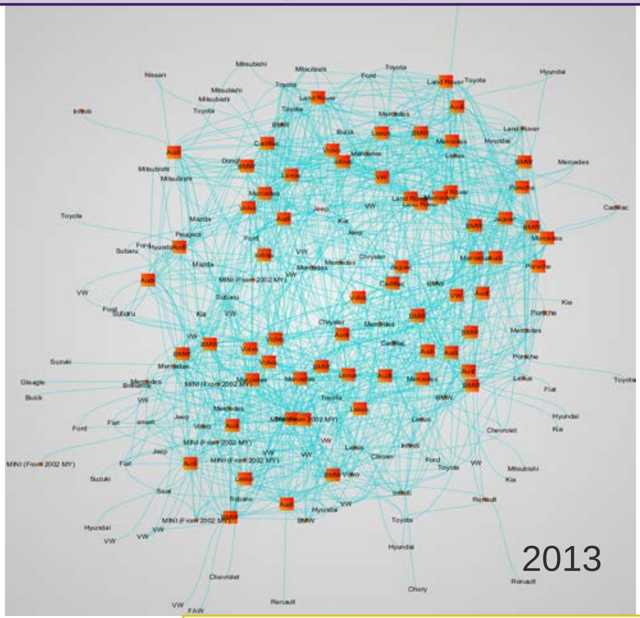
** $p < .01$; * $p < .05$

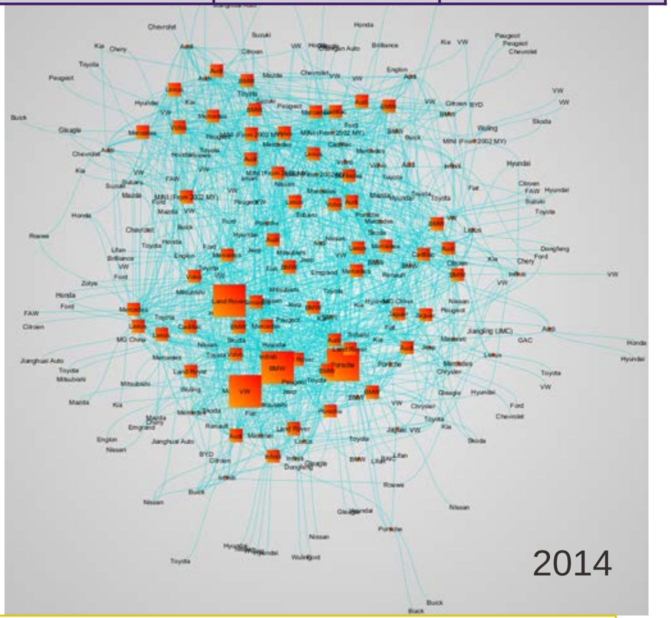
Key Insights

- Network effect: skewed market distribution
- Important factors: Price, fuel consumption, make origin, external styling, Turbo, AWD, autotrans
- Different processes during first stage and second stage manifested in the sign change of coefficients

On-going: Multi-year Network Evolvement

	Regular Vehicles (2013 | 2014)		Premium vehicles (2013 | 2014)	
Number of vehicles	289	302	100	101
Average degree (co-consideration)	23.63	22.48	28.94	33.33
Average cluster coefficient (three-way competition)	0.252	0.197	0.38	0.33





Observation 1: The size of network increases → more premium vehicles and more co-considerations on premium vehicles

Observation 2: Average cluster coefficient decreased → Three-way competition is less frequent in premium vehicles market in 2014 as compared to 2013.

On-going: Spatiotemporal Analytic Modeling for Customer Purchase

The incorporation of dependence in both time and space dimensions.

Time-Space Dynamic

$$y_t = \phi y_{t-1} + \gamma W_N y_{t-1} + \rho W_N y_t + X_t \beta + \epsilon_t$$

$N \times 1$ vector of cross section observations at time t; N is the number of provinces

Time autoregressive parameter; a constant but may evolve overtime

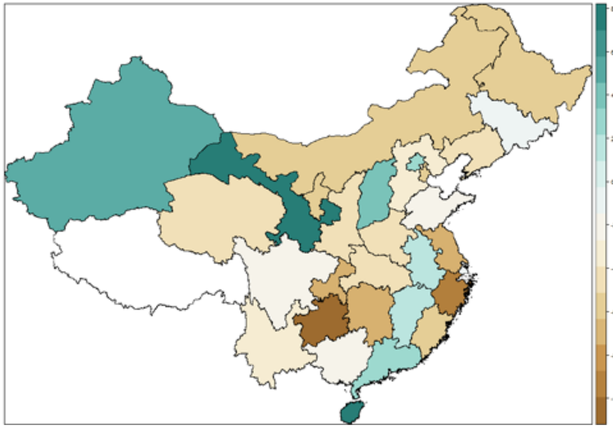
space-time autoregressive parameter

$N \times N$ spatial weights matrix W with N cross-sectional dimension

Spatial (contemporaneous) autoregressive parameter

$N \times K$ vector of Explanatory variables

$K \times 1$ vector of Regression coefficient



Future Use Cases:

- ✓ Forecast the sales for a particular province with the information of its neighboring locations in a previous period
- ✓ Use the latent space enhance the prediction.
- ✓ Use the most leading province to analyze how a particular car segment, e.g., EV/HV, propagates/diffuses throughout the market.

Thank You

NORTHWESTERN UNIVERSITY
QUAECUMQUE SUNT VERA
1851

NSF

Ford