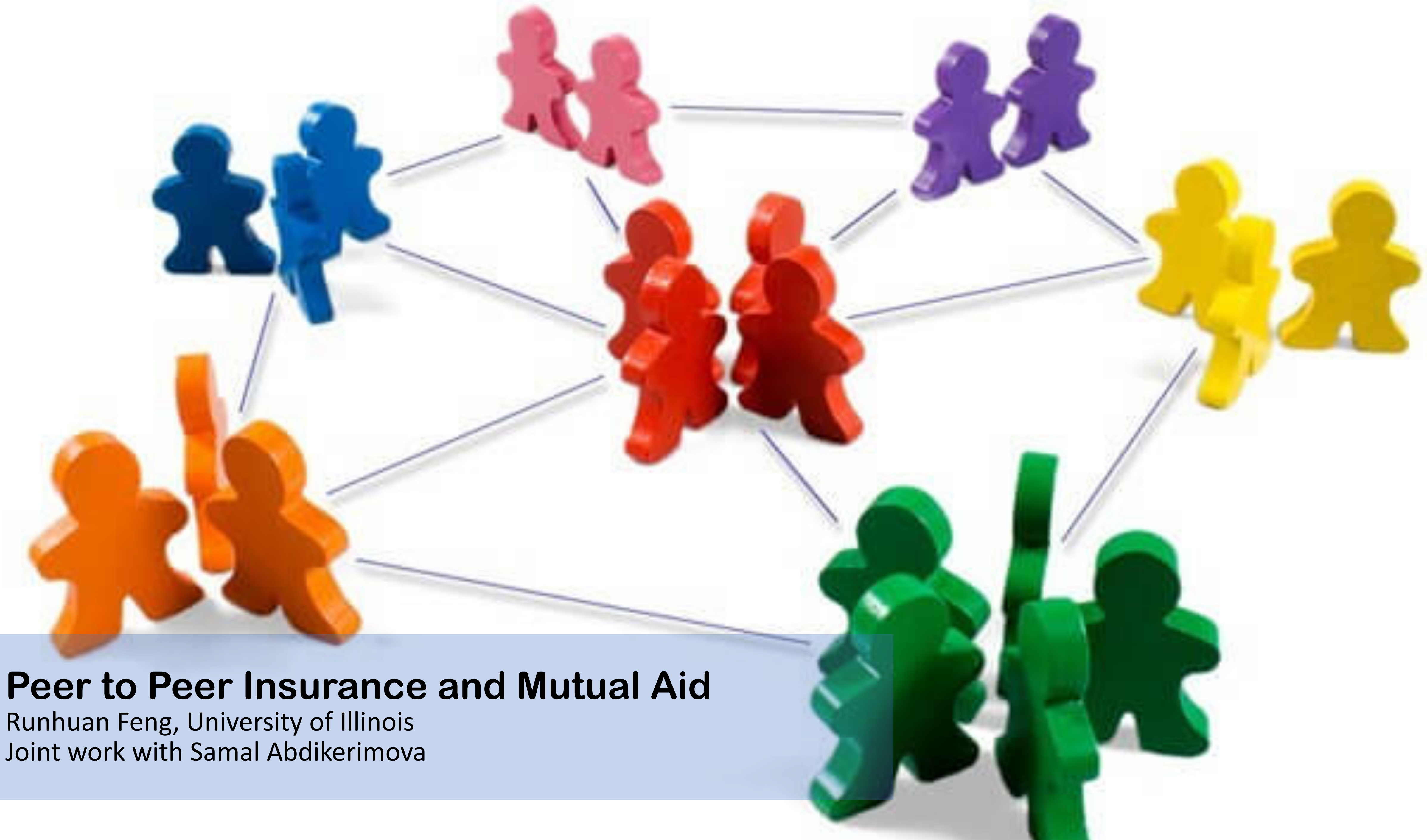
Leading Edge Property and Casualty Research in Academia

Runhuan Feng, University of Illinois at Urbana-Champaign

Peng Shi, University of Wisconsin- Madison

Emiliano Valdez, University of Connecticut



Peer to Peer Insurance and Mutual Aid

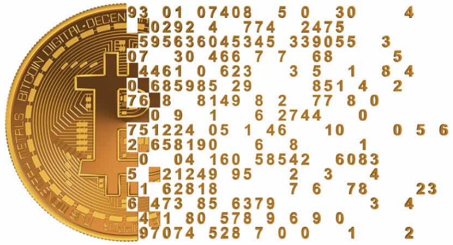
Runhuan Feng, University of Illinois
Joint work with Samal Abdikerimova

Decentralization /Disintermediation

- Technology
- Peer to Peer
- Sharing Economy

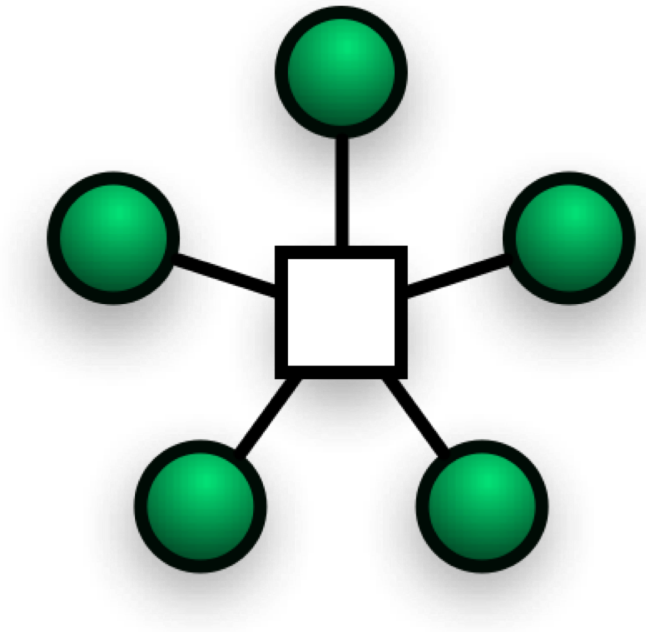
WIKIPEDIA
The Free Encyclopedia

Baidu百科
百度百科——全球最大中文百科全书

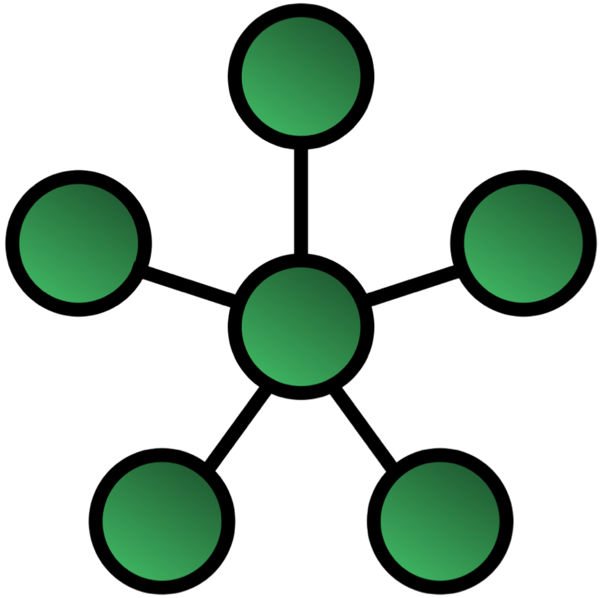


twitter

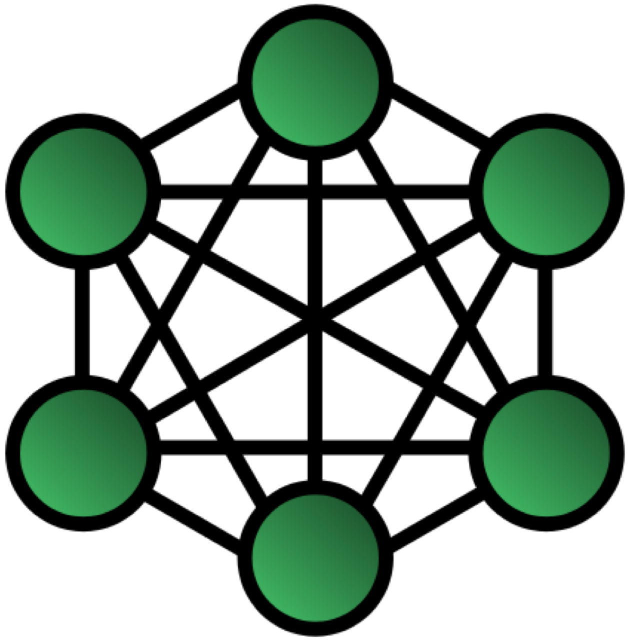
Historical Perspective



Primitive:
Mutual support



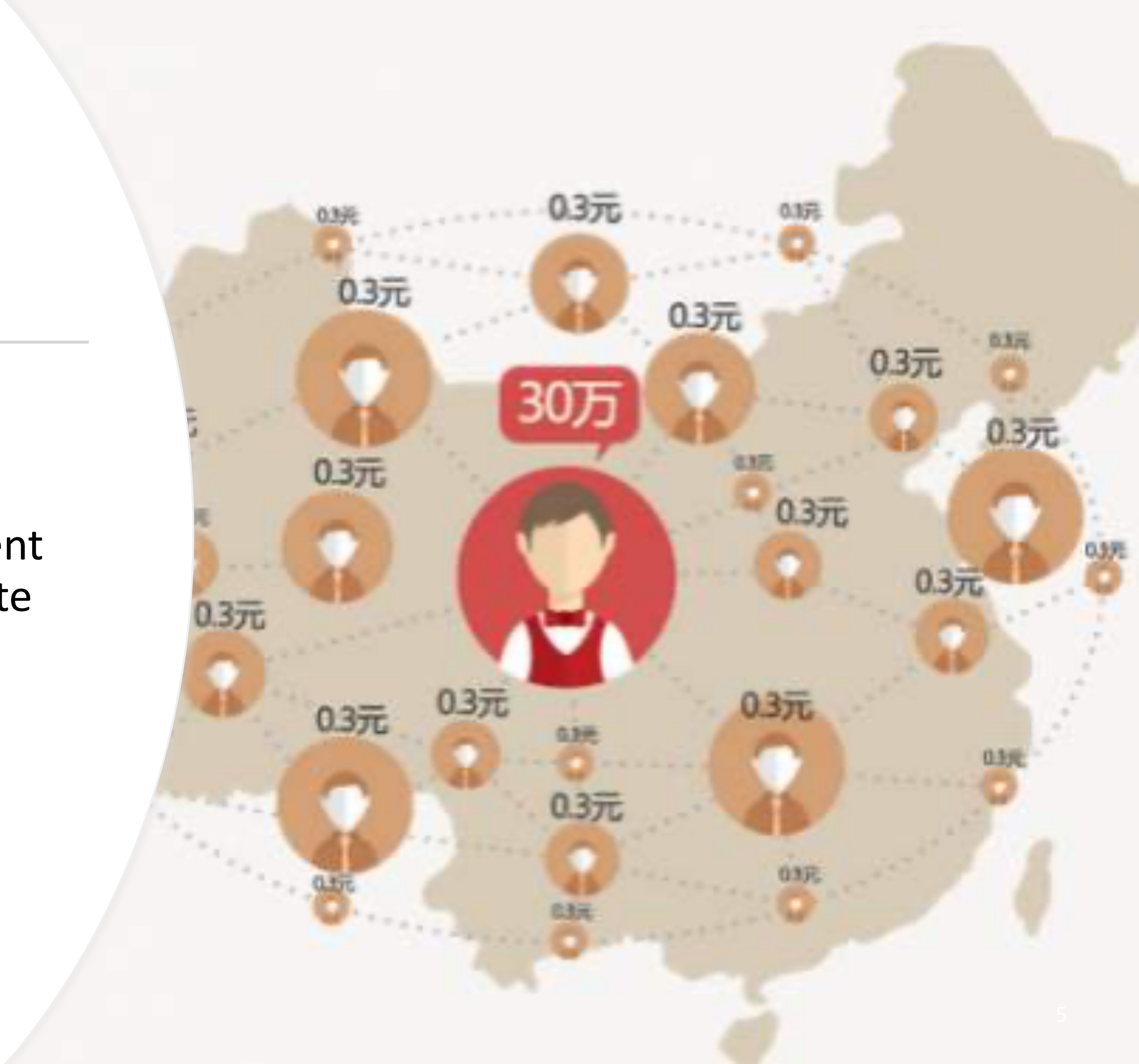
Standardization:
Modern insurance

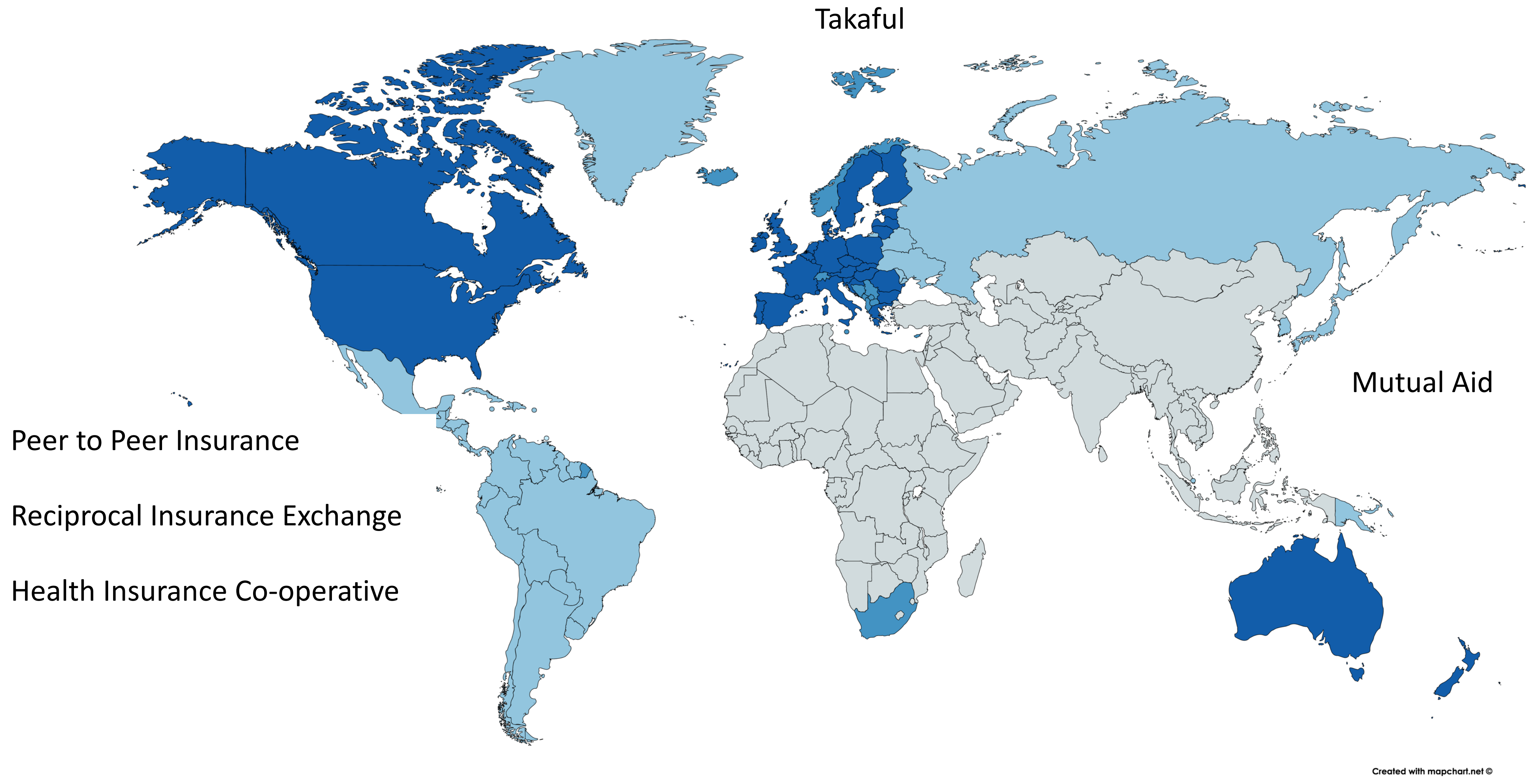


Peer to Peer:
Decentralized Insurance

Mutual Aid

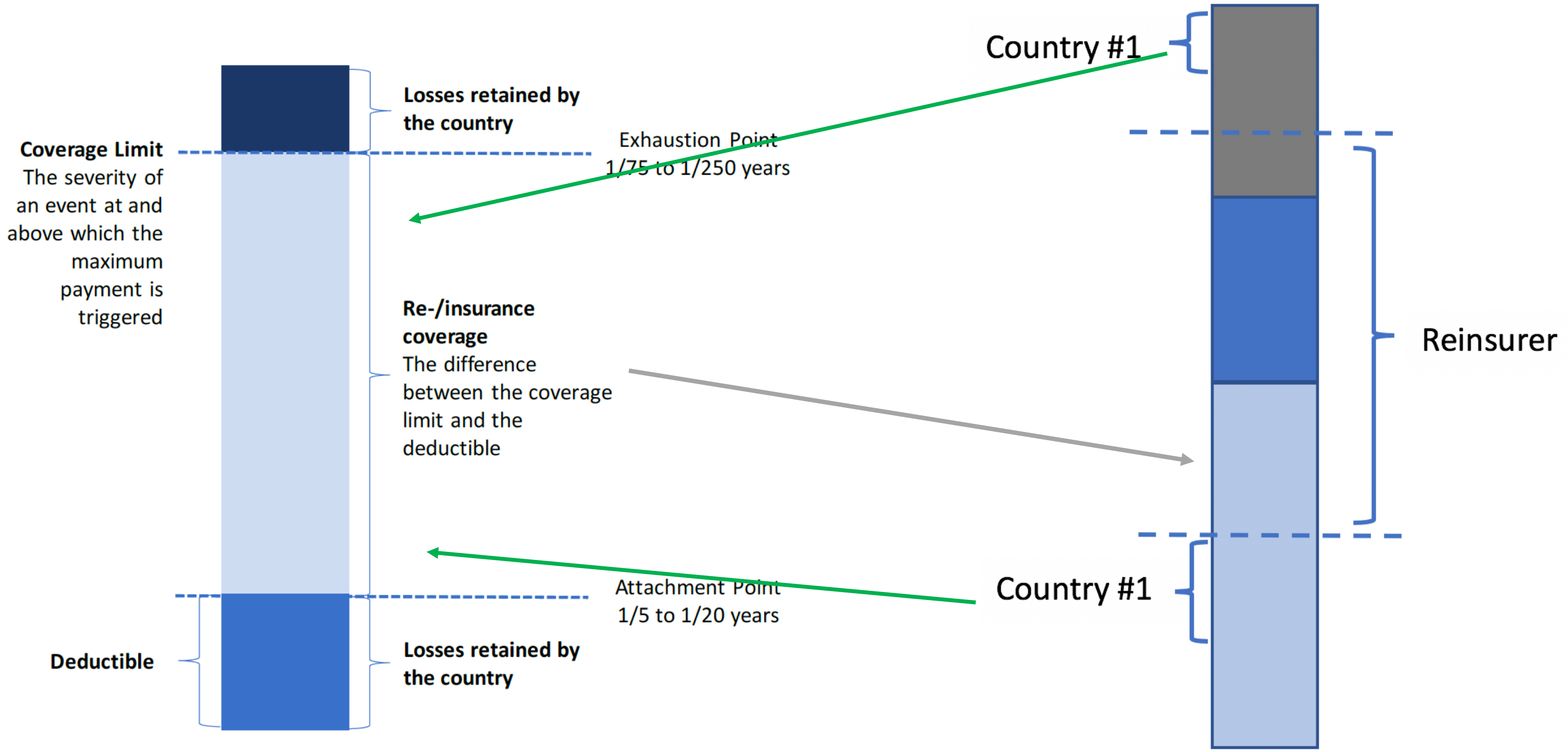
Mutual aid is a financial arrangement in a network of peers to compensate each other for losses





Takaful
Mutual Aid
Peer to Peer Insurance
Reciprocal Insurance Exchange
Health Insurance Co-operative
Created with mapchart.net ©

CAT Risk Pooling



Left panel is drawn from Bollman & Wang (2019)

Xianghubao

- Critical illness coverage

Age	Mild C.I.	Severe C.I.
30 days to 39 y.o.	50,000 RMB	300,000 RMB
40 to 59 y.o.	50,000 RMB	100,000 RMB

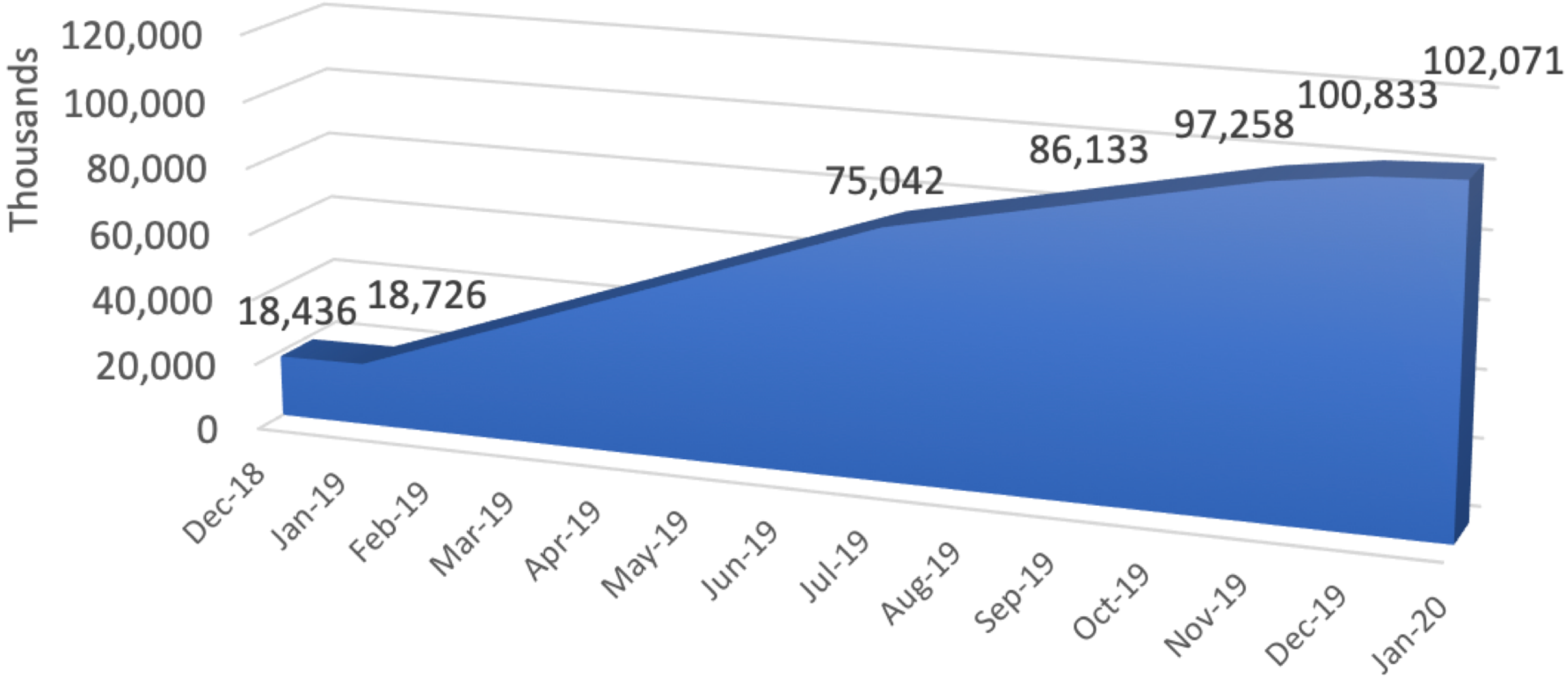
- Cost sharing

$$\text{Monthly cost} = \frac{\text{total mutual aids} + \text{management fee} - \text{remainder}}{\#\text{ of participants}}$$

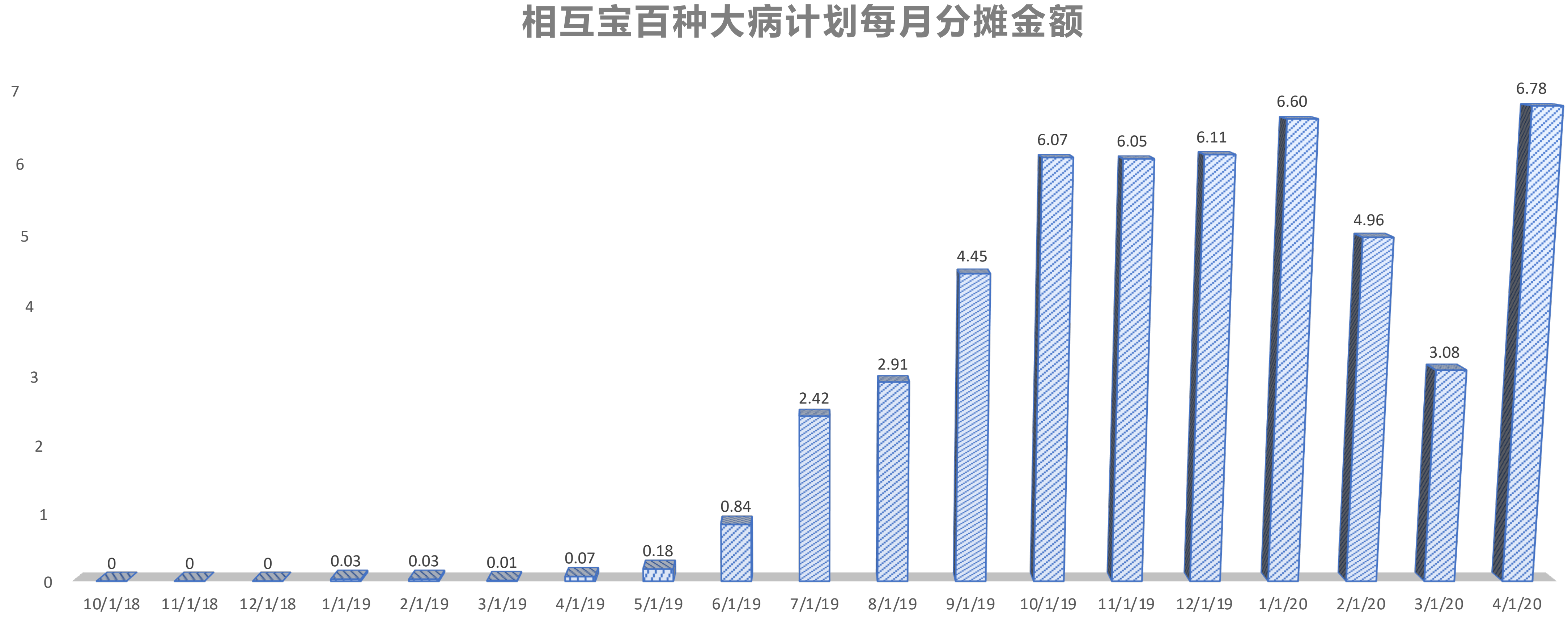
Example：Suppose there are 10 million users and 110 are diagnosed with mild CI. Management fee 8%

$$\text{Monthly } cost = \frac{5{,}500{,}000 \times (1+8\%) - 40{,}000}{10{,}000{,}000} = 0.59 \text{ RMB/person}$$

Growth of Xiaohubao Users

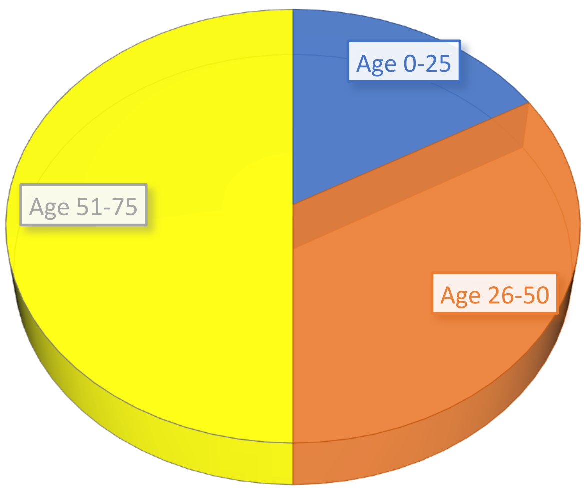


Monthly cost since established in 10/1/2018



- Adverse selection
- Differential pricing
- Two common methods
  - Equal cost sharing and differential benefit
  - Equal benefit and differential cost sharing

COST SHARING BY AGE GROUP



Online Mutual Aid （网络互助）

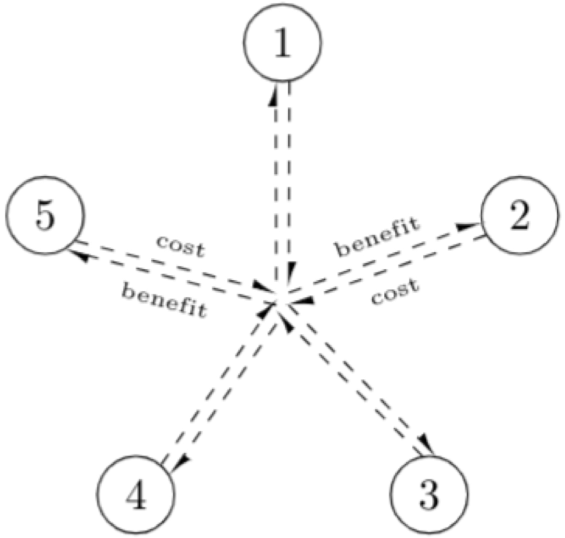
众托帮百万大病医疗互助计划

保障 111种 重大疾病和 77种 特定疾病

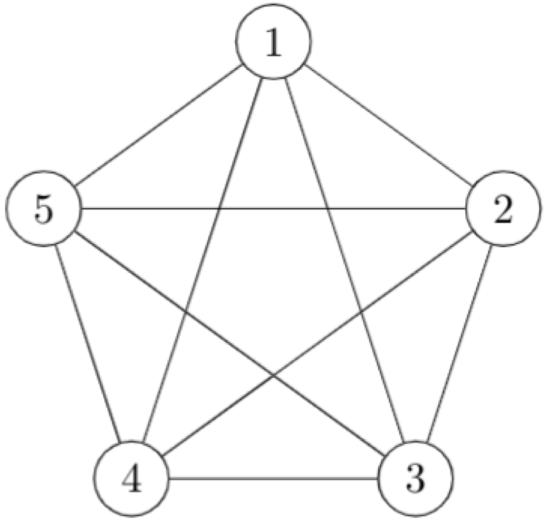
最高可获 50万元

会员年龄	0–25周岁	26–50周岁	51–75周岁
2016年	0	0	0
2017年	2.39	3.70	6.05
2018年	6.42	12.84	25.68
2019年	6.80	13.60	27.20
累计分摊(元)	15.61	30.14	58.93

×1 ×2 ×3



Peer-to-Peer Cost Allocation

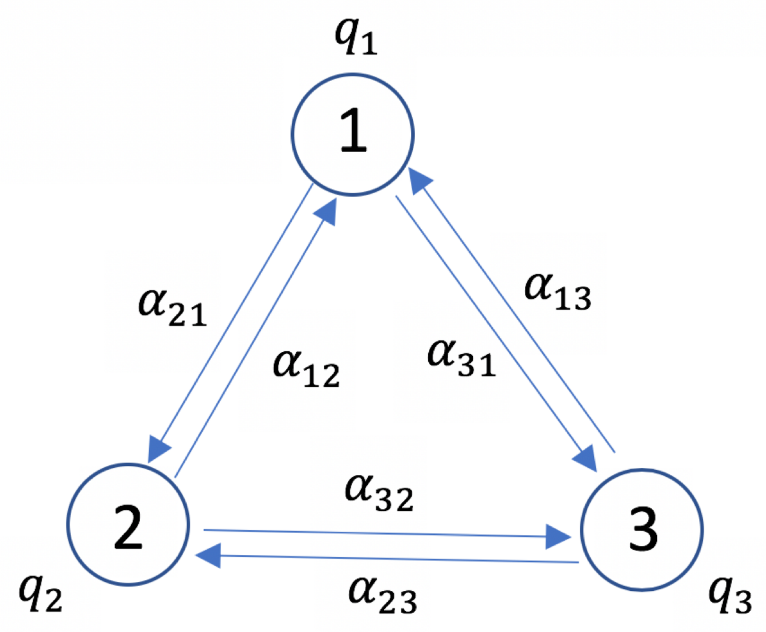


	Peer #1 (0-25)	**Peer #2 (26-50)**	**Peer #3 (51-75)**
Peer #1 (0-25)	1/6	1/6	1/6
Peer #2 (26-50)	1/3	1/3	1/3
Peer #3 (51-75)	1/2	1/2	1/2

	Peer #1	**Peer #2**	**Peer #3**
Peer #1	1/2	1/6	1/8
Peer #2	0	1/3	1/2
Peer #3	1/2	1/2	3/8

Peer-to-Peer Allocation Network

- Actuarial fairness
- Principle of indemnity
- Complete allocation
- Optimal designs



Limited fluctuation credibility

Minimum participants for <5% deviation with >90% chance

	Parameters	Case #1	Case #2	Case #3	Case #4	Case #5	Case #6
	s_1	10,000	3,000	10,000	10,000	10,000	10,000
	s_2	8,000	8,000	4,000	8,000	8000	8000
	q_1	0.090	0.090	0.090	0.200	0.090	0.090
	q_2	0.065	0.065	0.065	0.065	0.020	0.065
	n_2	10,000	10,000	10,000	10,000	10,000	15,000
(traditional insurance)	$\underline{n}_1^I$	10,945	10,945	10,945	4,330	10,945	10,945
	$\underline{n}_1^P$ (ICES)	5,557	3,929	7,147	2,920	8,974	5,107
	$\underline{n}_1^P$ (AAP)	668	107	820	737	910	430

ICES： intra-class equal split

AAP： altruistic allocation plan

Theoretical Analysis of Peer-to-Peer Risk Sharing

Conventional Insurance, P2P Insurance, Mutual Aid

S. Abdikerimova and R. Feng (2019) Peer-to-peer multi-risk insurance and mutual aid.

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3505646

Z. Chen, R. Feng, C. Liu and L. Wei (2020) Decentralized insurance and optimal risk pooling.

Assessing Hail Risk for Property Insurers

Peng Shi, University of Wisconsin-Madison

2020 CAS Annual Meeting

Joint work with

Glenn M. Fung, American Family Insurance
Daniel Dickinson, American Family Insurance

Outline

Hail is among the costliest thunderstorms:

- An individual hailstorm could result in losses of > 1 billion dollars
- In U.S. alone, property losses due to hail from 2005-2016 exceed $19 billion dollars

Hail damage to residential and commercial property tops the list of annual claims for most insurers in U.S.

- Average insured loss of around $850 million
- Higher than any other country in the world

Hail risk is difficult to insure:

- Hidden from property owners
- Volatile in frequency and severity
- Expanding its reach

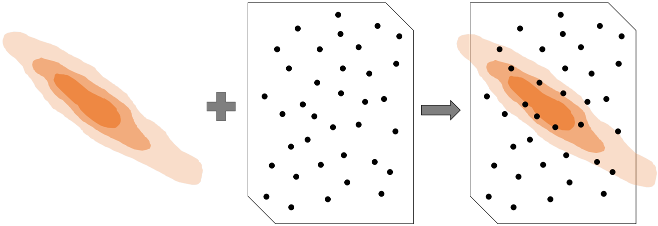
A pilot project collaborated with AmFam

We aim to provide a tool for insurers to assess hail risk from two aspects:

- Claim arrival patterns
- Financial impacts of hail risk

Two sources:

- the insurer's exposure that contains information on policies in-force
- radar data that provides information on hailstones in a hail swath



Data Collection

We focus on the hail storms in the state of Nebraska occurred between 2011 and 2015. There are in total 294 events during this period.

There are two levels of observations used in the analysis:

- Storm level: this allows us to examine the process of claims arrival
- Property level: this allows us to examine the financial impact

Claim Arrival

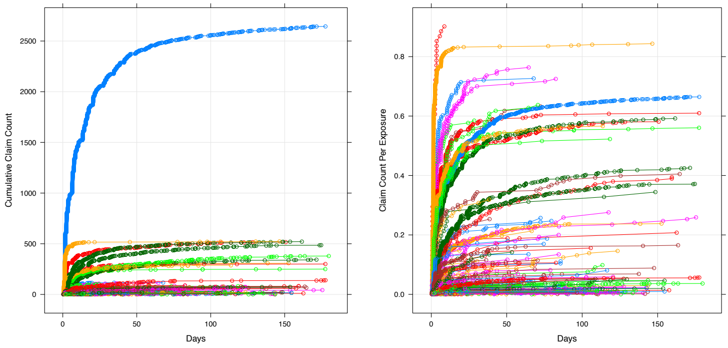


Figure: Cumulative event plots for insurance claims from hail storms. The left panel shows the number of claims and the right panel shows the claim count per exposure. Each curve corresponds to one hail storm.

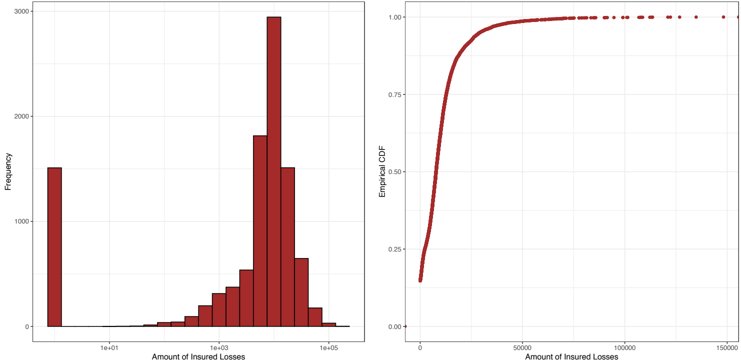


Figure: Distribution of insured amount of losses. The left panel shows the histogram and the right panel shows the empirical CDF.

Relationship

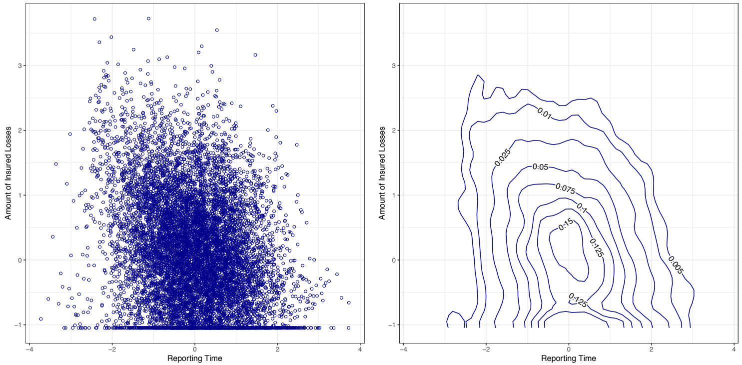


Figure: Relationship between reporting time and amount of insured losses. The left panel shows the scatter plot and the right panel shows the contour plot.

Claim Arrival Model

Claim arrival is modeled via a counting process:

- For storm i, the process of claims reporting is represented by $\{N_i(t), 0 \leq t\}$
- The counting process for claims arrival is specified via the marginal intensity function:

$$\lambda_i\left(t|H_i(t)\right) = \lim_{\Delta t \downarrow 0} \frac{\Pr\left(\Delta N_i(t) = 1|H_i(t)\right)}{\Delta t},$$

- To account for the storm-specific effects, we consider a random-effect Poisson model:

$$\lambda_i\left(t|H_i(t), V_i\right) = V_i \rho_i(t)$$

where $\rho_i(t) = \rho_0(t)\exp(x_i'\boldsymbol{\beta})$ and $V_i \sim Gamma(1/\theta, \theta)$.

Claim Severity Model

Claim amount is treated as marks of the process:

- Let Y_{ij} be the claim amount for claim reported at T_{ij}
- The joint distribution of (T_{ij}, Y_{ij}) is formulated using a copula C
- Because of the mass probability at zero, the conditional distribution of Y_{ij} given T_{ij} takes the form:

$$f_{Y|T}(y_{ij}|t_{ij}) = \begin{cases} c_1(F_T(t_{ij}), F_Y(y_{ij})) & y_{ij} = 0 \\ f_Y(y_{ij})c(F_T(t_{ij}), F_Y(y_{ij})) & y_{ij} > 0 \end{cases}$$

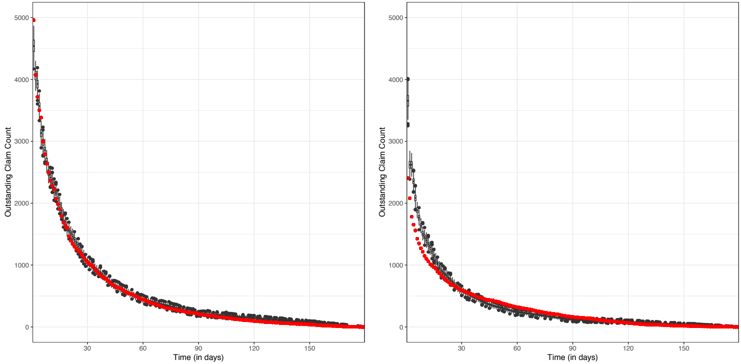


Figure: Dynamic predictive distribution of claim count. The left panel shows the prediction for the in-sample, and the right panel shows the prediction for the hold-out sample. the dots in the boxplots indicate the actual claim count.

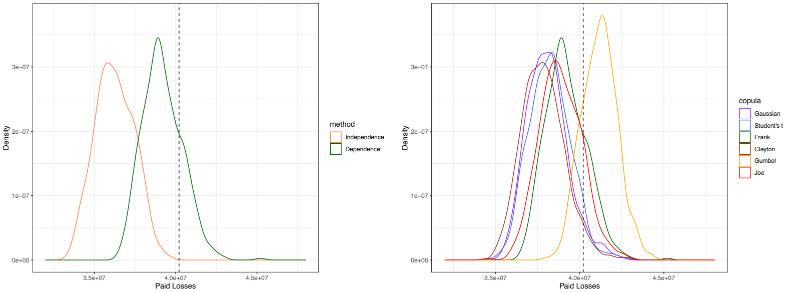


Figure: Predictive distribution of paid losses. The left panel compares the prediction between the independence and dependence models, and the right panel compares the prediction among various copula models. The vertical dotted line indicates the actual paid losses.

We employ a marked point process framework to model hail risk.

- Claim arrival time is correlated with the claim amount
- Claim amount distribution is complicated by risk retention

Thank you for your attention!

Cost-sensitive Multi-class AdaBoost for Understanding Driving Behavior with Telematics

Emiliano A. Valdez, PhD, FSA
joint work with Banghee So and Jean-Philippe Boucher

University of Connecticut and Université du Québec à Montréal (UQAM)

Introduction to telematics

- Telematics is about use of telecommunication devices and technology to transmit and store information.
  - wireless communication: plug-in device, already installed by manufacturers, mobile app
  - global positioning satellite (GPS), or global navigation satellite (GNS)
- Derived from the French word, télématique, which combines the words "telecommunications" and "computing science."
- Wide applications in the automobile industry.
- An intelligent device may be installed in the car to:
  - monitor and transmit driving information;
  - remotely control driverless cars;
  - monitor fleets for a systematic and efficient manner (e.g., Uber)

Telematics in auto insurance

- Introduced in early 2000's: Progressive Insurance, with General Motors.
- Primary purpose was to introduce Usage-based Insurance (UBI). Similar names floating:
  - Pay as you drive (PAYD), Pay how you drive (PHYD), Pay as you drive as you save (PAYDAYS), Pay per mile, Pay as you go (PASG)
- It enables insurers to collect driving metrics to enhance driver risk profile.
- Today: Progressive Snapshot, Traveler's IntelliDrive, AllState DriveWise, Esurance DriveSense, State Farm Drive Safe and Save.
- Economic benefits: studies show drivers save about 5-15% annually with UBI.
  - Social benefits e.g., reduces congestion, car emissions

Telematics data

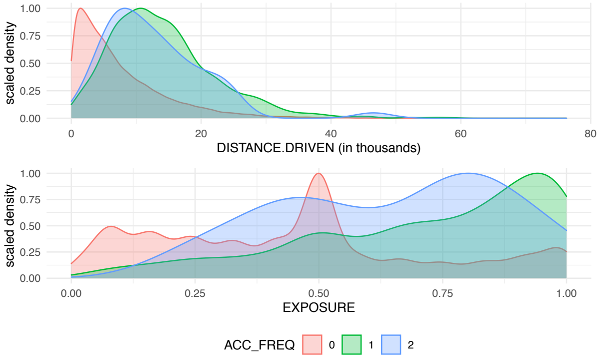
- Empirical data from Canadian-owned company offering insurance and investment products:
  - UBI auto program was first launched in year 2013
- Observation period: 2013–2016
- We have two possible response variables: number of accidents, number of claims
  - focus on accident frequency: a classification variable
- The sampled data on telematics were observations during the period for which 50,301 are used for training and another 21,574 for testing.
  - 97.1% with zero claims; 2.8% with exactly one claims; 0.1% with two+ claims

Traditional and telematics variables in the dataset

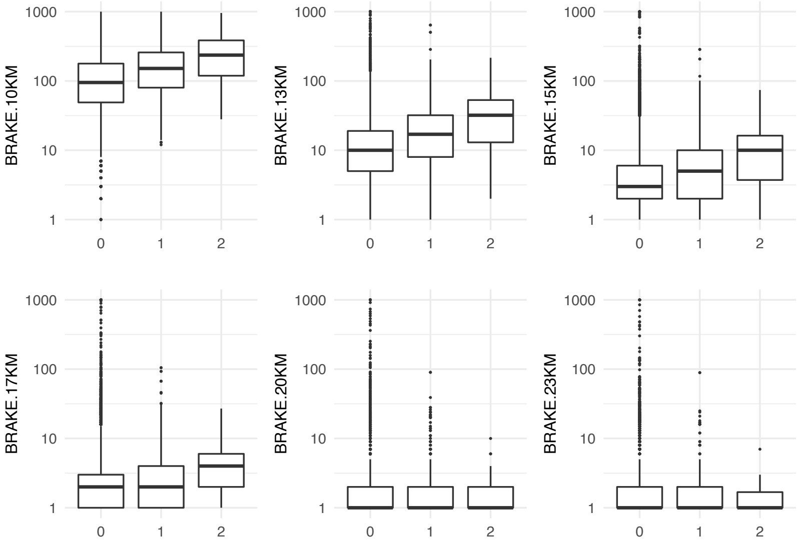
Type	Variable	Description
Traditional	DRIVER.AGE	Age of driver
	GENDER	Gender of the driver (M/F)
	VEHICLE.AGE	Vehicle age
	MARITAL	Marital status
	VEH.USE	Use of vehicle: Pleasure, Commute, Farmer, Business
	CREDIT.SCORE	Credit score of driver
	ZONE	Zone where driver lives: rural, urban
	ANN.KMS.DRV.SYSTEM	Kilometer driven declared by driver
	YRS.CLAIMS.FREE	Number of years claims free
	TERRITORY	Territory where vehicle is rated
Telematics	EXPOSURE	Exposure time in percentage of 365 days
	DISTANCE.DRIVEN	Total distance driven
	PCT.TRIP.xxx	Percent of driving day xxx of week: MON/TUE/.../SUN
	PCT.TRIP.xxx	Percent vehicle driven in xxx hrs: 2HRS/3HRS/4HRS
	PCT.xxx.DRIV	Percent vehicle driven in xxx of week: WKDAY/WKEND
	xx.RUSH.HOUR	Percent of driving in xx rush hours: AM/PM
	AVGDAY.USE.WKLY	Average number of days used per week
	ACCEL.xxKM	Number of sudden acceleration 10/13/15.../23 km/h/s per 1000km
	BRAKE.xxKM	Number of sudden brakes 10/13/15.../23 km/h/s per 1000km
	LTURN.EVENTxx	Number of left turn per 1000km with intensity 08/09/10/11/12
	RTURN.EVENTxx	Number of right turn per 1000km with intensity 08/09/10/11/12
Response	ACC_FREQ	Frequency of accidents during observation: $0/1/2+$

Accident frequency by DISTANCE.DRIVEN and EXPOSURE

Variable	Acc Freq	Count	Mean	Std Dev	Min	Q1	Median	Q3	Max
DISTANCE.DRIVEN	0	48822	7555.3	7149.4	0.1	2374.8	5395.7	10592.7	76271.8
	1	1430	14155.4	8257.3	253.9	8319.7	12657.4	18161.2	58759.2
	2	49	12834.9	7925.9	2247.8	7408.1	11408.3	16621.3	46527.4
EXPOSURE	0	48822	0.49	0.31	0.00	0.24	0.50	0.73	1.08
	1	1430	0.78	0.25	0.02	0.64	0.89	1.00	1.06
	2	49	0.74	0.26	0.23	0.50	0.80	1.00	1.06



Accident frequency by BRAKE events



Adaptive boosting (AdaBoost) algorithm

- AdaBoost combines several “weak learners” into a single “strong learner,” leading to improved predictions:
  - May be used for both classification and regression problems.
  - “Weak learners” in AdaBoost are decision trees with a single split (called decision stumps).
  - Algorithm is an iterative process, placing more weight on difficult to classify instances and less with those already classified well.
- Freund and Schapire (1996)
- AdaBoost does not work well for multi-class problems:
  - It requires the error of each weak learner to be better than by chance (i.e., 50%). Challenging with several classes.

Handling unbalanced classification data

- Unbalanced data: observed classes are not approximately equally represented
- Why is this a problem?
  - possible bias in predictions
  - gives a false sense of high accuracy
- Possible ways to handle:
  - Balance the classes: increase more observed minority and decrease observed majority
  - Random under-sampling: some observations discarded, lead to bias
  - Random over-sampling: no information loss but has tendency to overfit
- Increasingly popular technique is the use of SMOTE (Synthetic Minority Over-sampling Technique). Chawla, et al. (2002).
  - Take a feature vector, identify its nearest neighbor (e.g. KNN), locate a new observation in between. Repeat process

Competitive models considered in our work

Compare methods or algorithms that combine the benefits of resampling and boosting including:

- `SAMME` (Stagewise Additive Modeling using Multi-class Exponential loss function);
  - The exponential loss function is convex, exponentially increases with negative values, which makes it more sensitive to outliers.
- `SAMME` with `SMOTE` sampling;
- `SMOTEBoost`, described in Chawla et al. (2003), is an approach for learning from minority classes based on a combination of `SMOTE` and `AdaBoost.M2`; and
- `RUSBoost`, described in Seiffert et al. (2010), is an algorithm that has the same goal as `SMOTEBoost` but replaces `SMOTE` sampling with random undersampling.

SAMME: multi-class AdaBoost

Initial sample weights $D_1(i) = \frac{1}{N}, \quad i = 1, \ldots, N$

Get weak classifier $h_t : X \to k \in \{1, \ldots, K\}$

Train weak classifier using the distribution D_t, for $t = 1, \ldots, T$

Train sub-classifier h_t and compute

$$\epsilon_t = \frac{\sum_{i=1}^{N} D_t(i) I(y_i \neq h_t(\boldsymbol{x}_i))}{\sum_{i=1}^{N} D_t(i)}$$

Calculate weight of the t-th sub-classifier

$$\alpha_t = \log\left(\frac{1-\epsilon_t}{\epsilon_t}\right) + \log(K-1)$$

Update sample weights

$$D_{t+1}(i) = \frac{D_t(i) \exp(-\alpha_t I(y_i = h_t(\boldsymbol{x}_i)))}{\sum_{j=1}^{N} D_t(j) \exp(-\alpha_t I(y_j = h_t(\boldsymbol{x}_j)))}$$

Return (final) classifier

$$H(\boldsymbol{x}_i) = \underset{k}{\operatorname{argmax}} \sum_{t=1}^{T} \alpha_t I(h_t(\boldsymbol{x}_i) = k)$$

`Ada.C2`: cost-sensitive binary AdaBoost

Initial sample weights $D_1(i) = \frac{1}{N}, \quad i = 1, \ldots, N$

Get weak classifier $h_t : X \to k \in \{1, \ldots, K\}$

Train weak classifier using the distribution D_t, for $t = 1, \ldots, T$

Train sub-classifier h_t and compute

$$\epsilon_t = \frac{\sum_{i=1}^{N} C(y_i)\, D_t(i) I(y_i \neq h_t(\boldsymbol{x}_i))}{\sum_{i=1}^{N} C(y_i)\, D_t(i)}$$

Calculate weight of the t-th sub-classifier

$$\alpha_t = \frac{1}{2} \log\left(\frac{1-\epsilon_t}{\epsilon_t}\right)$$

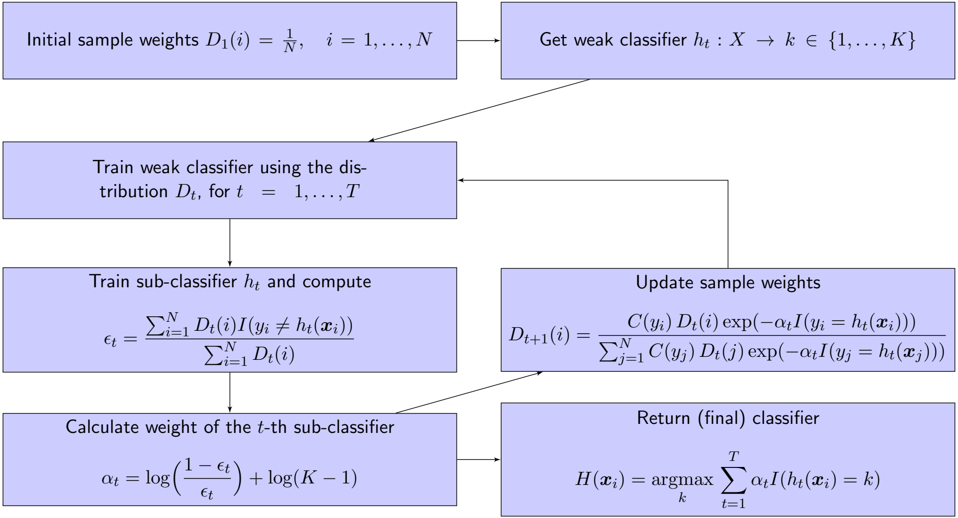
Update sample weights

$$D_{t+1}(i) = \frac{C(y_i)\, D_t(i) \exp(-\alpha_t I(y_i = h_t(\boldsymbol{x}_i)))}{\sum_{j=1}^{N} C(y_j)\, D_t(j) \exp(-\alpha_t I(y_j = h_t(\boldsymbol{x}_j)))}$$

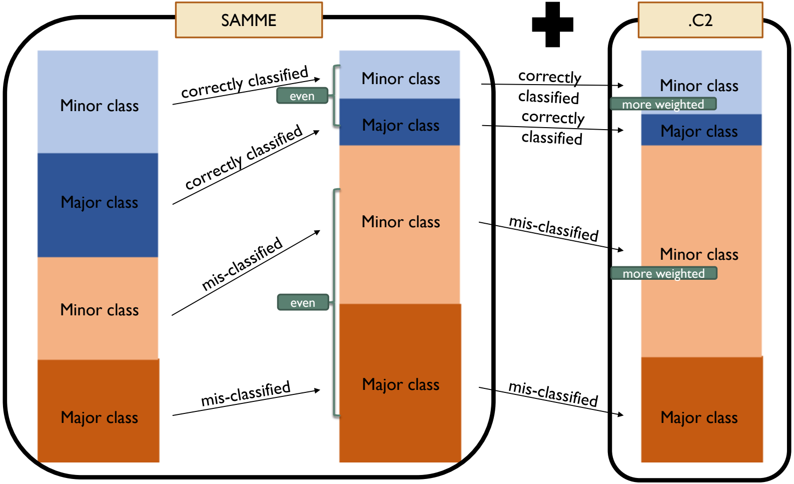
Return (final) classifier

$$H(\boldsymbol{x}_i) = \operatorname*{argmax}_k \sum_{t=1}^{T} \alpha_t I(h_t(\boldsymbol{x}_i) = k)$$

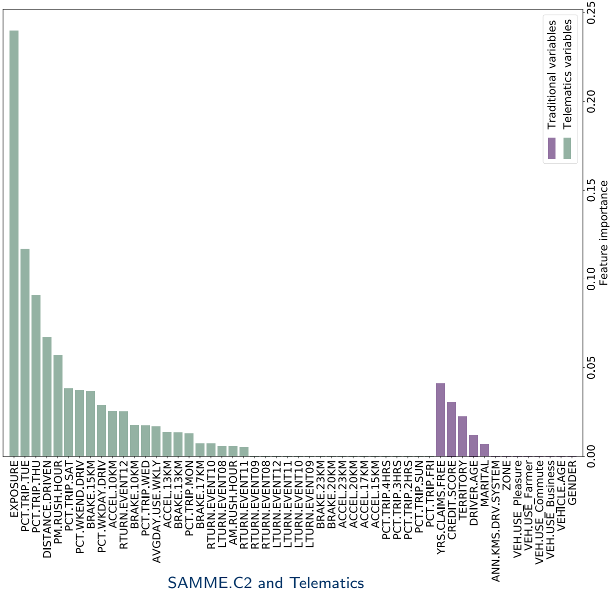
SAMME.C2: cost-sensitive multi-class AdaBoost



Visualizing the effect of the SAMME.C2 algorithm on classifying majority/minority class



Feature importance based on the `SAMME.C2` predictions



Confusion tables based on the model fit of `SAMME.C2`

With traditional variables only

Actual

Predicted	Acc 0	Acc 1	Acc 2^+	Tot row
Acc 0	11674	175	1	11850
Acc 1	5319	274	3	5596
Acc 2^+	3908	201	19	4128
Tot col	20901	650	23	21574

With traditional and telematics variables

Actual

Predicted	Acc 0	Acc 1	Acc 2^+	Tot row
Acc 0	14553	108	0	14661
Acc 1	5669	420	5	6049
Acc 2^+	679	122	18	819
Tot col	20901	650	23	21574

Performance using Macro Average Geometric of recall statistics (MAvG)

Accident Frequency	Recall statistics				
	SAMME	SAMME with SMOTE	RUSBoost	SMOTEBoost	SAMME.C2
Accident 0	1.00	0.84	0.68	0.99	0.70
Accident 1	0.06	0.38	0.46	0.18	0.65
Accident 2^+	0.00	0.35	0.04	0.13	0.78
MAvG	0.02	0.48	0.24	0.28	0.71

Recall statistics, also called sensitivity, for class i, R_i, is proportion of observations in class i correctly classified. We aggregate these recall statistics using the geometric average and define

$$\text{MAvG} = (R_1 \times R_2 \times \cdots \times R_K)^{1/K}.$$

A better performing classifier is one that gives a larger value of MAvG.

Predicted classification based on the model fit of SAMME.C2 for hypothetical drivers varying telematics information but keeping same values of traditional variables: DRIVER.AGE=40, VEHICLE.AGE=10, GENDER=Female, MARITAL=Married, VEHICLE.USE=Commute, and ZONE=Rural

Hypothetical driver	EXPOSURE	PCT.TRIP.TUE	PCT.TRIP.THU	DISTANCE.DRIVEN	PM.RUSH.HOUR	YRS.CLMS.FREE	PCT.TRIP.SAT	PCTWKEND.DRIV	BRAKE.15KM	CRED.SCORE	PCTWKDAY.DRIV	ACCEL.10KM	RTURN.EVENT12	Predicted ACC_FREQ
H1	0.3	0.10	0.10	5000	0.10	30	0.10	0.50	10	850	0.50	30	10	0
H2	0.8	0.13	0.13	15000	0.15	15	0.20	0.20	20	650	0.80	70	20	1
H3	0.8	0.18	0.18	15000	0.20	10	0.08	0.80	150	650	0.20	150	150	2^+

References

- Fernandez, A. et al. (2018). *Learning from Imbalanced Data Sets*. Springer: Switzerland.
- Hastie, T. et al. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer: New York.
- Zhu, J. et al. (2009). Multi-class AdaBoost. *Statistics and Its Interface*. 2: 349-360.
- Chawla, N.V. et al. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*. 16: 321-357.
- So, B., Boucher, J.P., and Valdez, E.A. (2020). Cost-sensitive Multi-class AdaBoost for Understanding Driving Behavior with Telematics. Available at arXiv: `https://arxiv.org/pdf/2007.03100.pdf`