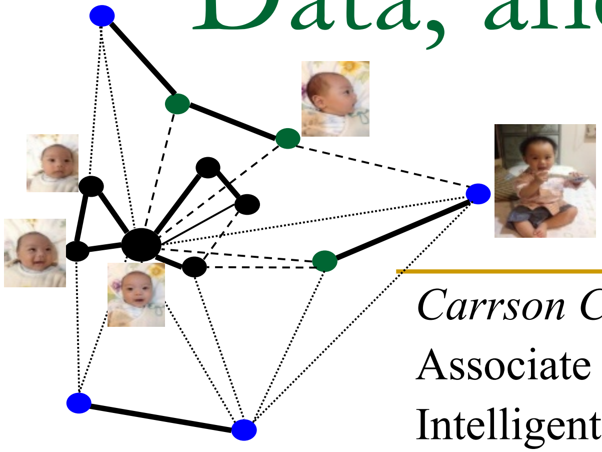
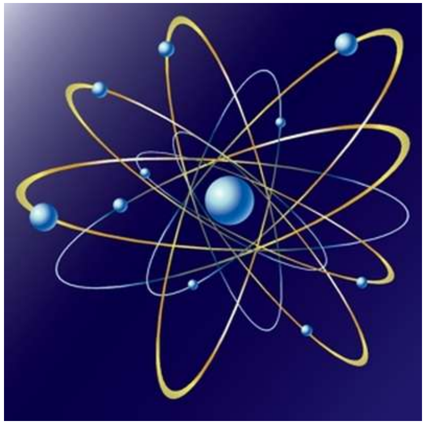
Graph Signal Processing For Big Data, and How I Got Here



Carrson C. Fung
Associate Professor
Intelligent Modeling and Optimal Design Group (IMOD)
Communication Electronics and Signal Processing Lab (CommLab)
Institute of Electronics
National Chiao Tung University

Who Am I Professionally?



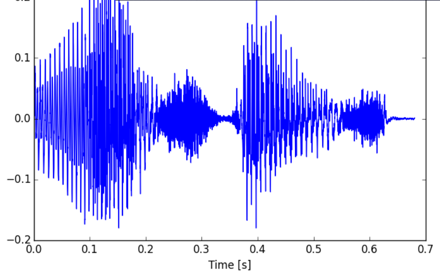
MPEG

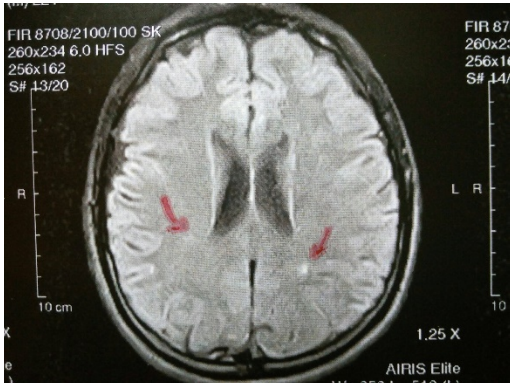
PERCEPTUAL AUDIO CODER

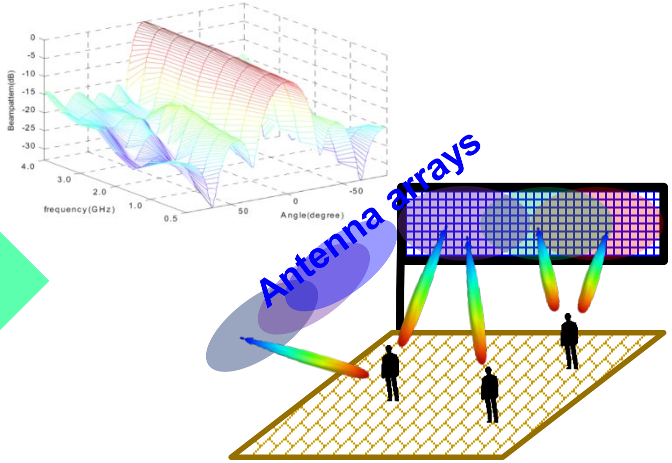
INTRODUCING

Speech Recognition

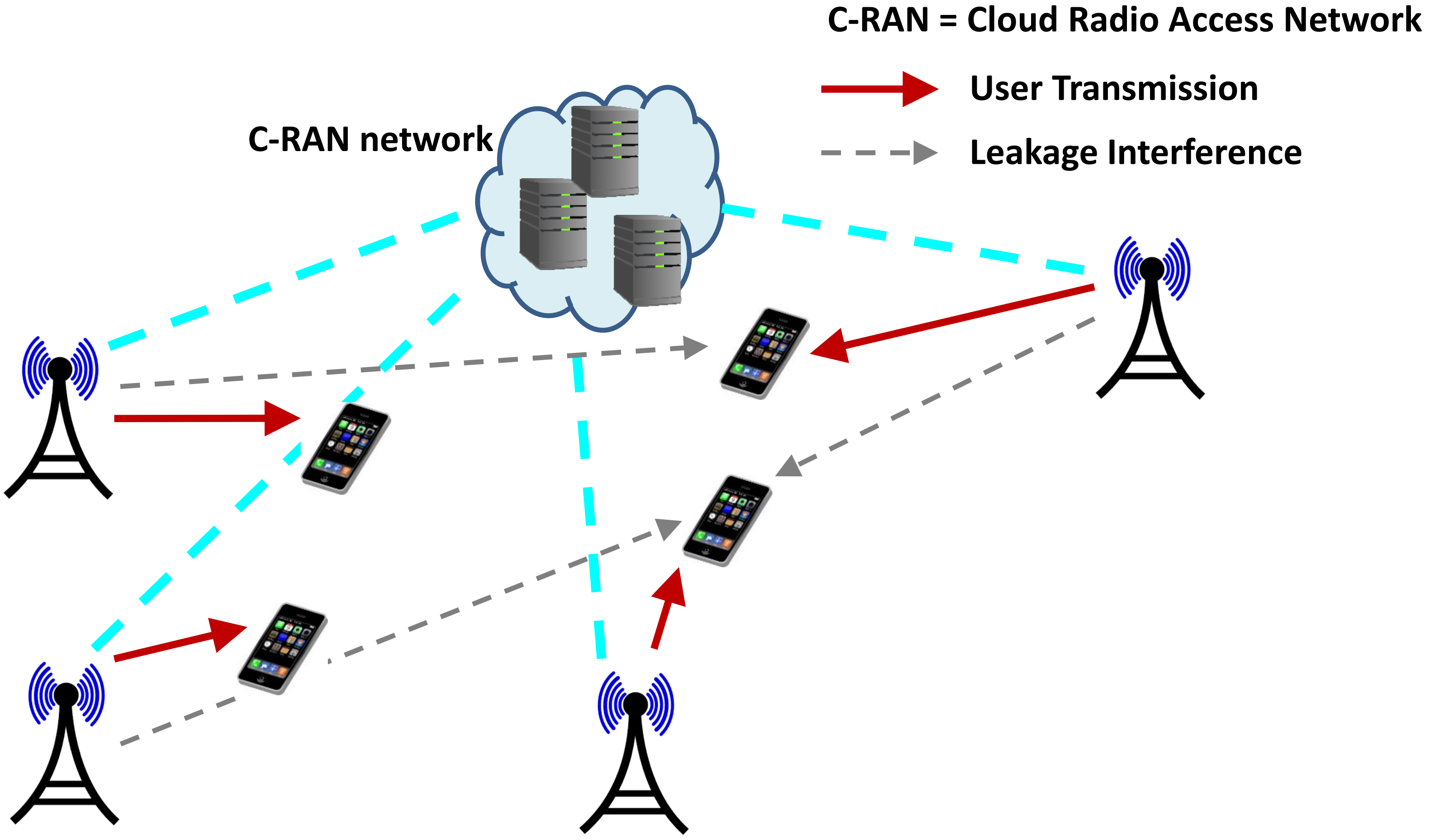
Convert speech to text and analyze its intent during any voice call.







Traditional Cellular Network

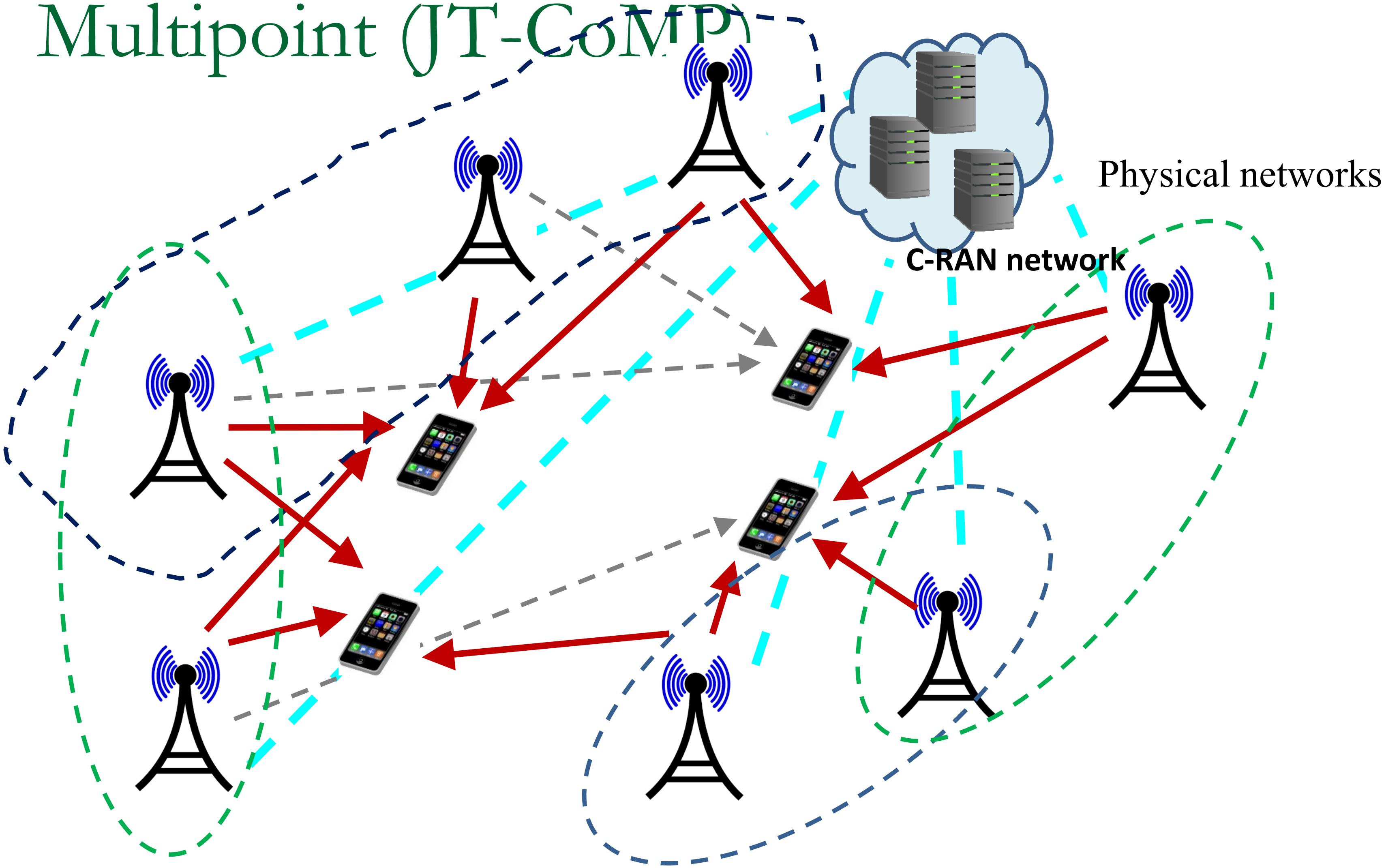


C-RAN = Cloud Radio Access Network

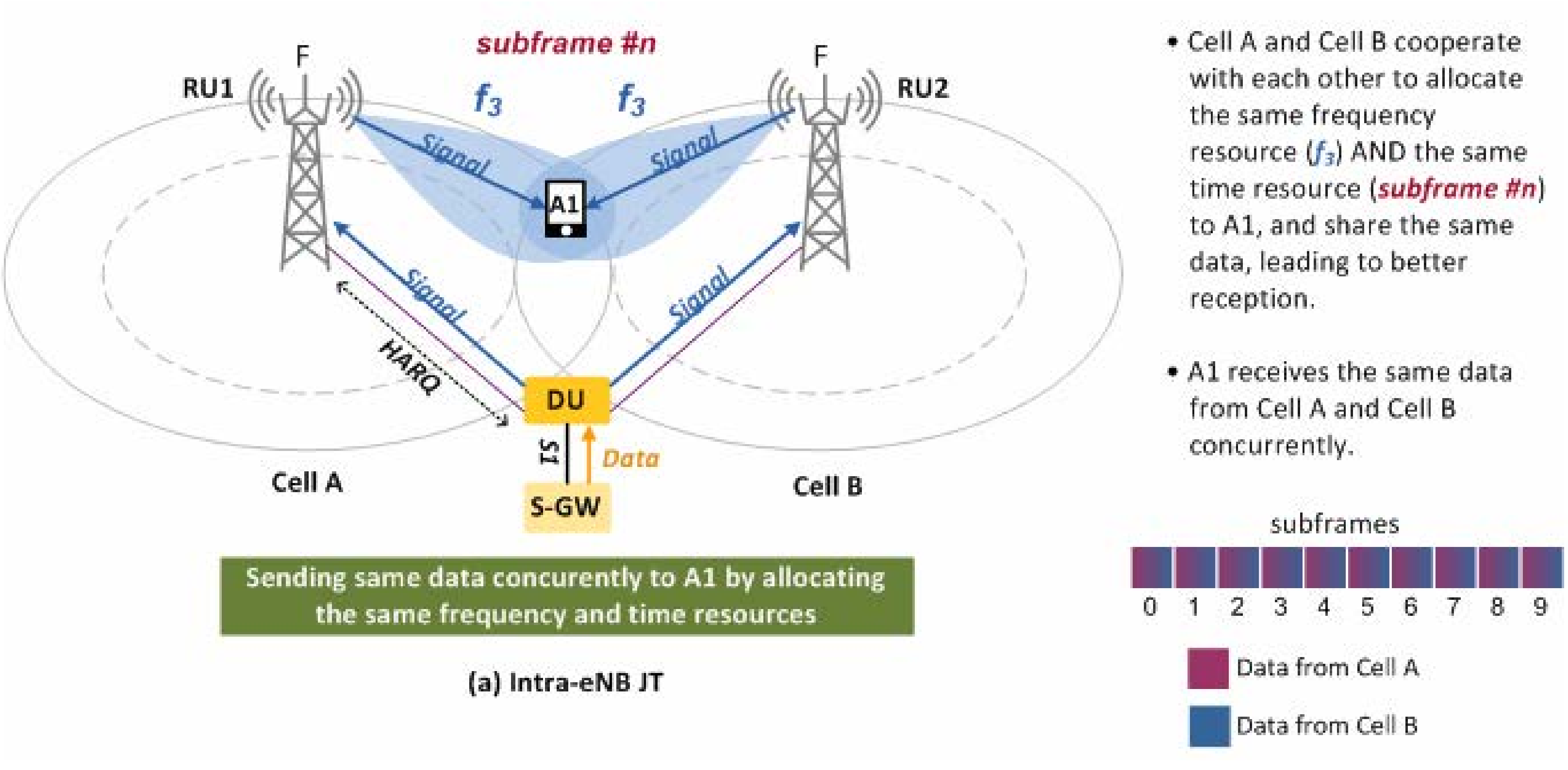
→ User Transmission

⇢ Leakage Interference

Joint Transmission-Coordinated Multipoint (JT-CoMP)



JT-CoMP



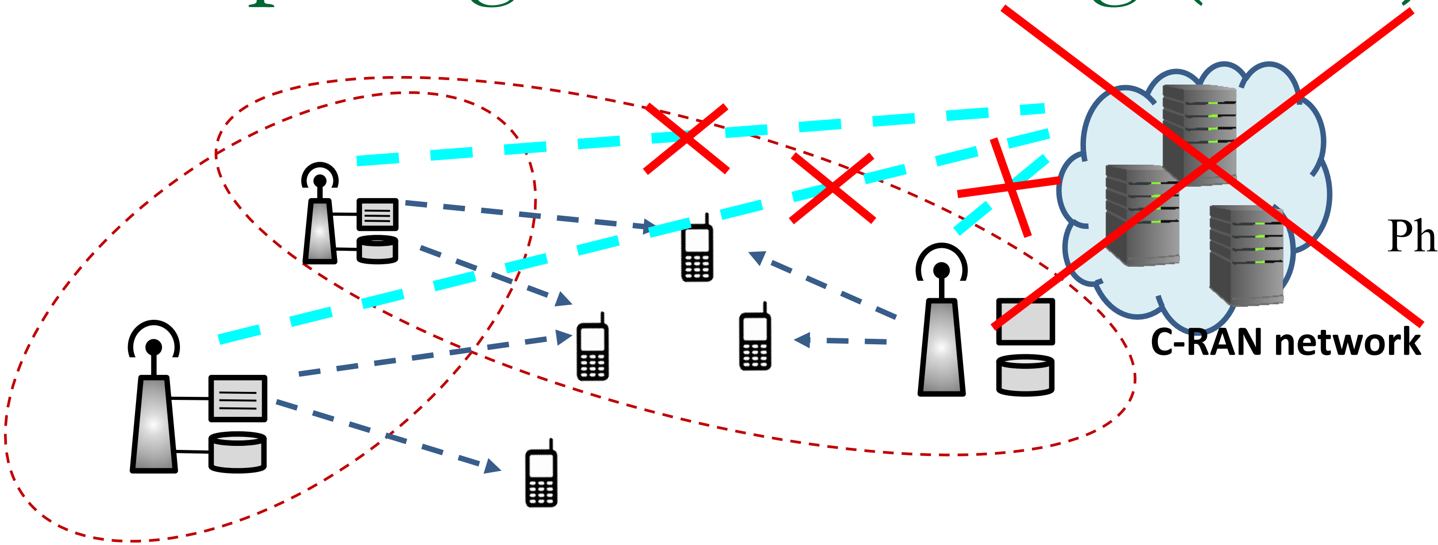
Sending same data concurently to A1 by allocating the same frequency and time resources

(a) Intra-eNB JT

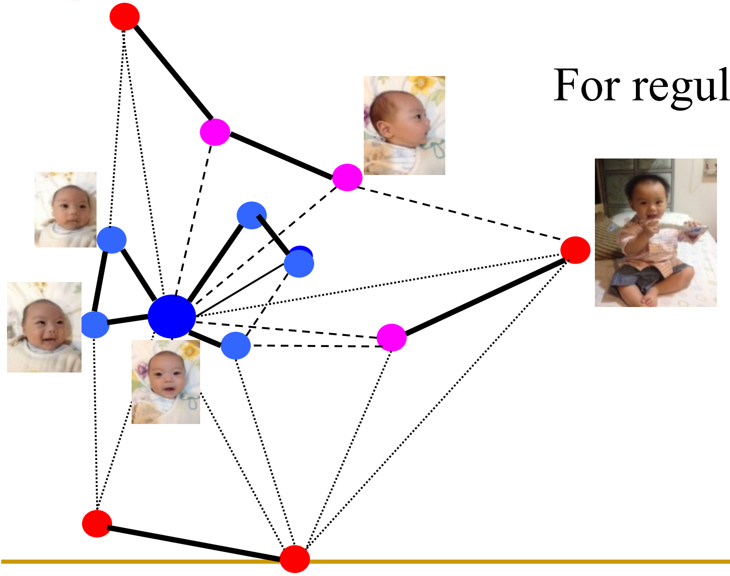
- Cell A and Cell B cooperate with each other to allocate the same frequency resource (f_3) AND the same time resource (***subframe #n***) to A1, and share the same data, leading to better reception.
- A1 receives the same data from Cell A and Cell B concurrently.

🟥	Data from Cell A
🟦	Data from Cell B

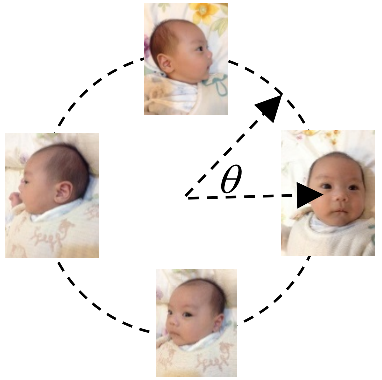
Graph Signal Processing (GSP)



Physical networks



For regular signals

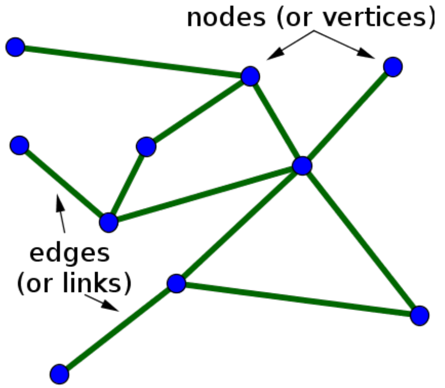


For information networks

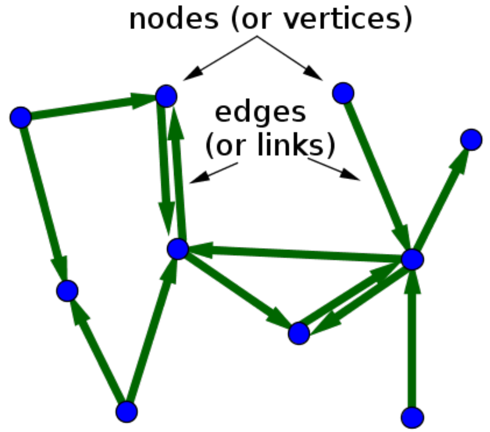
What is a Graph?

- Collection of vertices and edges
  - A set of *nodes* and the *relationships* that connect them
  - $G(V,E)$
- Represents entities as nodes
- And how these entities relates to the world as relationship

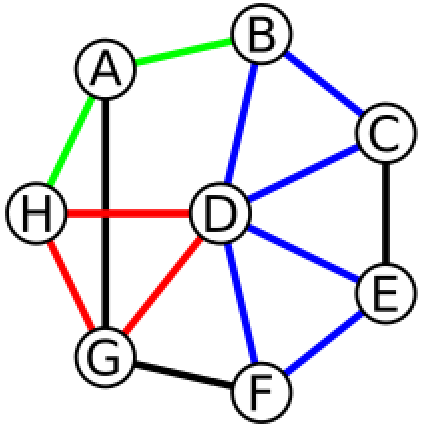
Graph Types



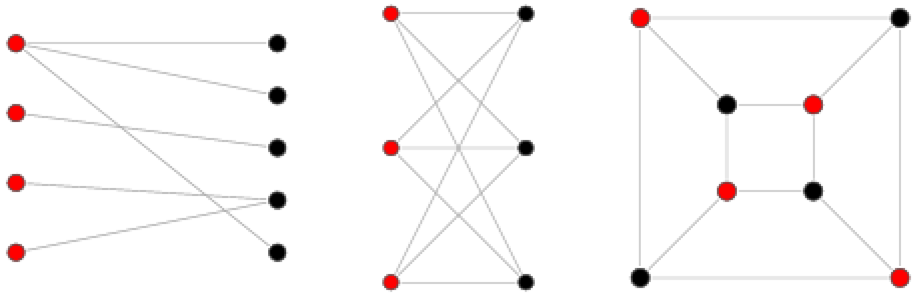
Undirected



Directed



Cyclic



Bipartite

Misconceptions about Big Data (Data Science)

- It’s only multimedia
- It means you need to deal with lots of data
  - 5 V’s: volume (curse of dimensionality), velocity, variety (data and knowledge fusion), veracity (statistics 101), value
- It’s the same as machine and deep learning (AI)
  - Domain knowledge
  - Data and knowledge fusion/cross-space fusion
  - Data mining
  - Machine and deep learning
  - Ingenuity about which model to use for data (science & art)

Danger in Big Data

- Danger in using data science approach (WMD) [O2016]
  - Opacity, scale, damage, lack of feedback (p.109)
    - Example: university ranking
    - Algorithm used lacks transparency
    - Affects many applicants
    - Overemphasis on alumni donation, SAT scores, student-teacher ratios, acceptance rates, fund raising
    - E.g. Colleges game the system by increasing fund raising ➔ campus improvement ➔ student acceptance increased ➔ ranking increased
      - Tuition amount not in algorithm and this can drive away good students
- Good example: FICO (Fair, Issac, and Co.)
  - Founded in 1956 to provide credit score
  - Evaluate risk that an individual will default on a loan
    - Only looks at the borrower's finances: debt loan, bill-paying record
    - Color-blind
    - Transparent and always updated based on your action

C. O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Crown Publishing, New York, 2016.

Which aspects of data science do we deal with (so far)?

- Volume and velocity
  - Data collected from everywhere
    - Difficult/impossible to perform centralized processing
      - ➔ Distributed algorithm
- Veracity
  - Cross-layer design
- Variety
  - Data fusion not enough
    - ➔ Knowledge fusion is more appropriate
- Value
  - If no value, why would we do it?

Which aspects of data science do we deal with (so far)?

- Applications
  - Communications
    - Exploiting cloud AND edge/Fog computing resources
      - ➔ Graph model + distributed consensus algorithm
    - Predictive/preemptive analysis of communications network performance
      - ➔ Graph model + statistical data model + machine learning + fusion
  - Autonomous vehicles (land/area/sea)
    - Data association and object tracking (track fusion)
      - ➔ Graph model + distributed consensus algorithm

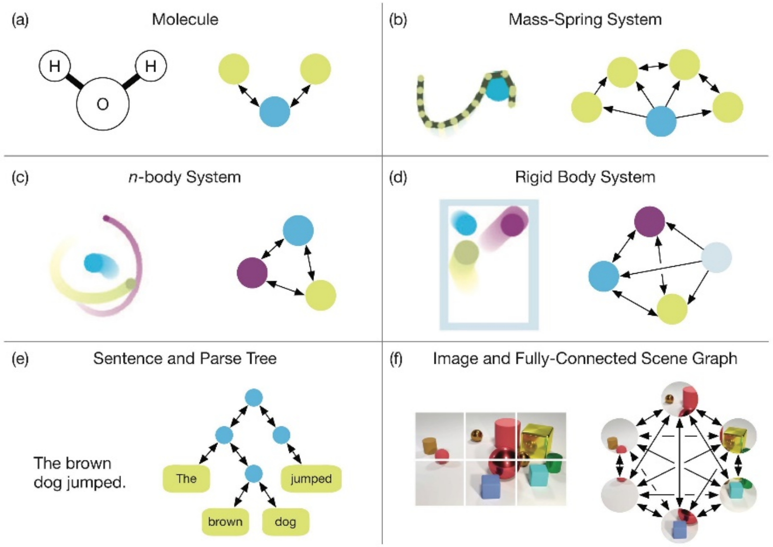
Why Signal Processing Matters

- ***Data science** is the study of the generalizable extraction of knowledge from data, yet the key word is science. It incorporates varying elements and builds on techniques and theories from many fields, including **signal processing**, **mathematics** **(optimization)**, probability models, machine learning, deep learning, computer programming, data engineering, pattern recognition, visualization, uncertainty modeling, data warehousing and high performance computing with the goal of extracting meaning from data and creating data products. Data Science is not restricted to only big data, although the fact that data is scaling up makes big data an important aspect of data science.*

T. Singh, "Why data scientists are crucial for AI transformation," *https://www.forbes.com/sites/cognitiveworld/2018/09/13/why-data-scientists-are-crucial-for-ai-transformation/amp/*, Sep. 13, 2018.

Why Graphs?

- Human cognition makes the strong assumption that the world is composed of objects and relations
  - A lot of things requires relationship to describe
- Training deep learning models with unstructured data fails to capture relationship
  - Graph network [B2018]
- Surprisingly, hard to capture relationships even relational database
  - Use graph database instead



P.W. Battaglia *et al.*, "Relational inductive biases, deep learning, and graph network," *arXiv:1806.01261*, Oct. 2018.

GDSP Basics: Delay

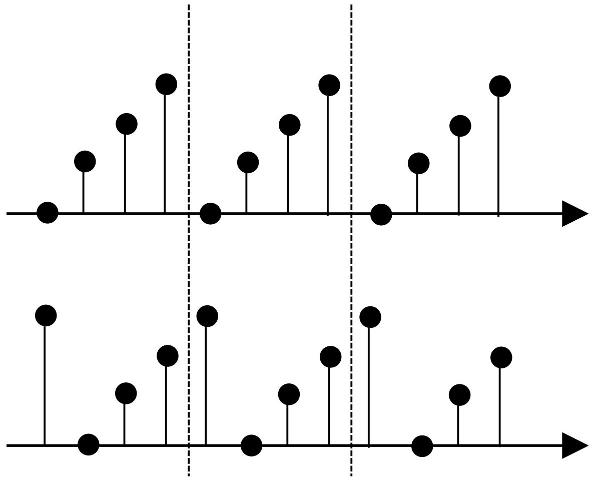
$$\mathbf{s} : \mathcal{V} \to \mathcal{C}$$

$$v_n \mapsto s_n$$

DSP

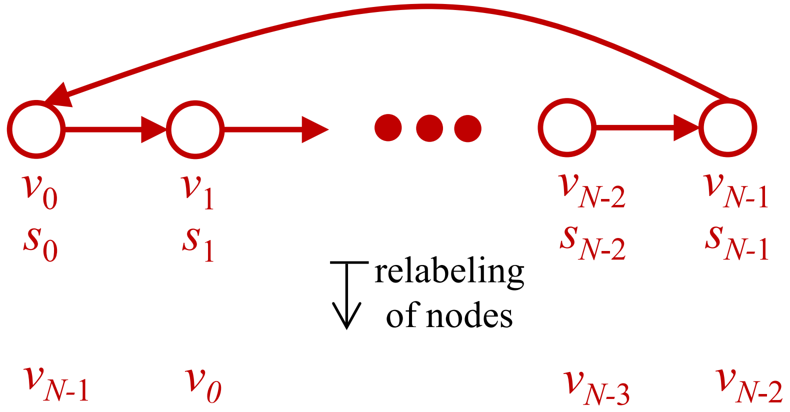
(shift)

$$s[n] \longrightarrow \boxed{z^{-1}} \longrightarrow s[(n\text{-}1)_N]$$



GSP

(graph shift)



$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 1 & 0 \end{bmatrix}$$

$$\hat{\mathbf{s}} = \mathbf{A}\mathbf{s}$$

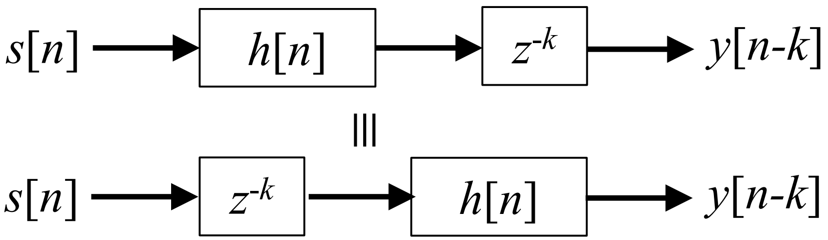
GSP Basics

- GSP in **A** equivalent to z^{-1} in DSP
- **A**: adjacency matrix
  - Describes relationship between nodes
  - $[\mathbf{A}]_{ij} = w_{ij}$
    - Models directed graph
  - A symmetric when $w_{ij} = w_{ji}$
    - Modes undirected graph
- $[\mathbf{D}]_{ii} = \sum_{j=1}^{N} [\mathbf{A}]_{ij}, \ [\mathbf{D}]_{ij} = 0, \text{ for } i \neq j$
- $\mathbf{L} = \mathbf{D}\text{-}\mathbf{A}$ is called the Laplacian

GSP Basics: Shift Invariance and Filtering

DSP

(time invariance)



GSP

(shift invariance)

$$\mathbf{A}(\mathbf{Hs}) = \mathbf{H}(\mathbf{As})$$

(filter)

$$h(x) = h_0 + h_1 x + \cdots + h_L x^L$$

(graph filter)

$$\mathbf{H} = h(\mathbf{A}) = h_0\mathbf{I} + h_1\mathbf{A} + \cdots + h_L\mathbf{A}^L$$

A. Sandryhaila and J.M.F. Moura, ``Discrete signal processing on graphs," *IEEE Trans. on Signal Processing*, vol. 61(7), pp. 1644-1656, Apr. 2013.

DSP Basics: Fourier Basis for LTI Systems

$$s[n] = e^{j\omega n} \longrightarrow \boxed{\text{LTI}, h[n]} \longrightarrow y[n] = H(e^{j\omega})\, e^{j\omega n} = H(e^{j\omega})\, s[n]$$

Suppose $L_h = 3$ and $L_s = 4$:

$$y[n] = h[n] * s[n] = \begin{bmatrix} y[0] \\ y[1] \\ y[2] \\ y[3] \\ y[4] \\ y[5] \end{bmatrix} = \begin{bmatrix} h[0] & 0 & 0 & 0 \\ h[1] & h[0] & 0 & 0 \\ h[2] & h[1] & h[0] & 0 \\ 0 & h[2] & h[1] & h[0] \\ 0 & 0 & h[2] & h[1] \\ 0 & 0 & 0 & h[2] \end{bmatrix} \begin{bmatrix} s[0] \\ s[1] \\ s[2] \\ s[3] \end{bmatrix} \Leftrightarrow \mathbf{y} = \mathbf{Hs}$$

$$y[n] = H(e^{j\omega})\, s[n] \Leftrightarrow \mathbf{y} = H(e^{j\omega})\, \mathbf{s}$$

$$\boxed{\mathbf{y} = \mathbf{Hs} = H(e^{j\omega})\, \mathbf{s}}$$

The Fourier basis $e^{j\omega}$ is eigenvector (eigenfunction) for **all** LTI systems

Graph Fourier Transform Basis

$$\mathbf{H} = h(\mathbf{A}) = h_0\mathbf{I} + h_1\mathbf{A} + \cdots + h_L\mathbf{A}^L$$

$$\begin{aligned}
\mathbf{H} &= h\left(\mathbf{A}\right) = h\left(\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1}\right) \\
&= \sum_{\ell=0}^{L-1} h_\ell \left(\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1}\right)^\ell = \sum_{\ell=0}^{L-1} h_\ell \mathbf{V}\left(\boldsymbol{\Lambda}\right)^\ell \mathbf{V}^{-1} \\
&= \mathbf{V}h\left(\boldsymbol{\Lambda}\right)\mathbf{V}^{-1}, \qquad h\left(\boldsymbol{\Lambda}\right) = \mathrm{Diag}\left(\left[h\left(\lambda_0\right) \ \cdots \ h\left(\lambda_{N-1}\right)\right]^T\right)
\end{aligned}$$

$$\boxed{\begin{aligned}
\mathbf{H}\mathbf{v}_\ell &= \mathbf{V}h(\boldsymbol{\Lambda})\mathbf{V}^{-1}\mathbf{v}_\ell \\
&= \mathbf{V}h(\boldsymbol{\Lambda})\,\mathbf{e}_\ell \\
&= h(\lambda_\ell)\mathbf{v}_\ell
\end{aligned}}$$

The eigenvcctors of **A** are eigenvectors of **all** LSI graph systems

GSP Fourier Transform

$$\mathbf{A}=\begin{bmatrix}0 & 0 & 0 & \cdots & 0 & 1\\ 1 & 0 & 0 & \cdots & 0 & 0\\ 0 & 1 & 0 & \cdots & 0 & 0\\ \vdots & \vdots & \ddots & \ddots & \ddots & 0\\ 0 & 0 & \cdots & 1 & 0 & 0\\ 0 & 0 & \cdots & 0 & 1 & 0\end{bmatrix}=\mathbf{W}_N\begin{bmatrix}e^{j\frac{2\pi 0}{N}} & & \\ & \ddots & \\ & & e^{j\frac{2\pi(N-1)}{N}}\end{bmatrix}\mathbf{W}_N^H, \qquad \left[\mathbf{W}_N^H\right]_{nk}=\frac{1}{\sqrt{N}}e^{-j\frac{2\pi}{N}nk} \quad \text{(DFT matrix)}$$

$$\mathbf{V}^{-1}=\mathbf{W}_N^H$$

Columns of **V** are graph spectral components, and eigenvalues of **A**, i.e. $e^{j2\pi n/N}$, for n=0, 1, …, N-1, are the graph frequencies

Filtering Graph Signals

DSP

$$\tilde{\mathbf{y}} = \tilde{\mathbf{H}}\mathbf{s} \Leftrightarrow \tilde{\mathbf{y}} = \mathbf{W}_N \underbrace{\begin{bmatrix} H[0] & & \\ & \ddots & \\ & & H[N-1] \end{bmatrix} \underbrace{\mathbf{W}_N^H \mathbf{s}}_{\text{FT of signal } s[n]}}_{\text{Pointwise multiplication with } H[k]}$$

$$H[k] = H\left(e^{j\frac{2\pi k}{N}}\right), \text{ for } k = 0,1,\ldots,N-1$$

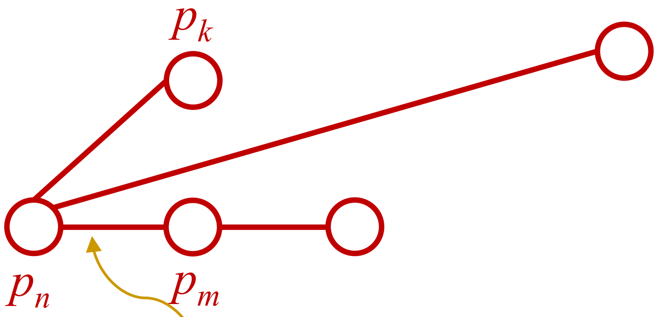
GSP

$$\mathbf{H} = h(\mathbf{A}) = \mathbf{V}h(\mathbf{\Lambda})\mathbf{V}^{-1},$$

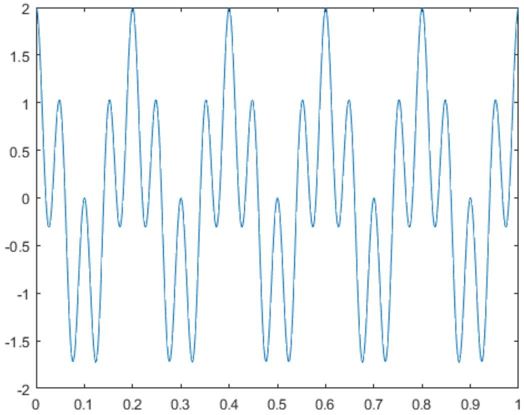
$$h(\mathbf{\Lambda}) = \mathrm{Diag}\left(\left[h(\lambda_0) \ \cdots \ h(\lambda_{N-1})\right]^T\right)$$

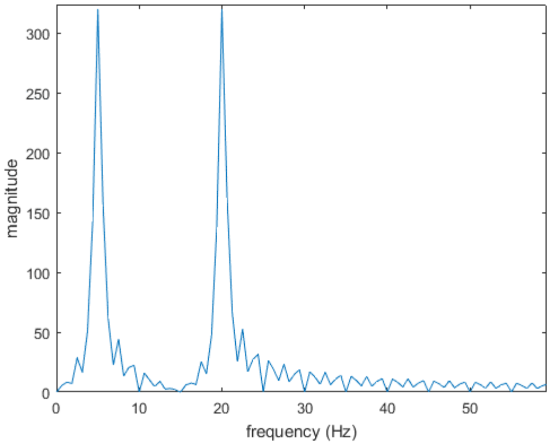
$$\begin{aligned}\tilde{\mathbf{y}} = \tilde{\mathbf{H}}\mathbf{s} \Leftrightarrow \tilde{\mathbf{y}} &= \mathbf{V}h(\mathbf{\Lambda})\mathbf{V}^{-1}\mathbf{s} = \mathbf{V}h(\mathbf{\Lambda})\mathbf{s}_T \\ &= \mathbf{V}\,\underbrace{\mathrm{Diag}\left(\left[h(\lambda_0)[\mathbf{s}_T]_0 \ \cdots \ h(\lambda_{N-1})[\mathbf{s}_T]_{N-1}\right]^T\right)}_{\text{Pointwise multiplication in (graph) frequency domain}}\end{aligned}$$

Application: Image Segmentation (Clustering)



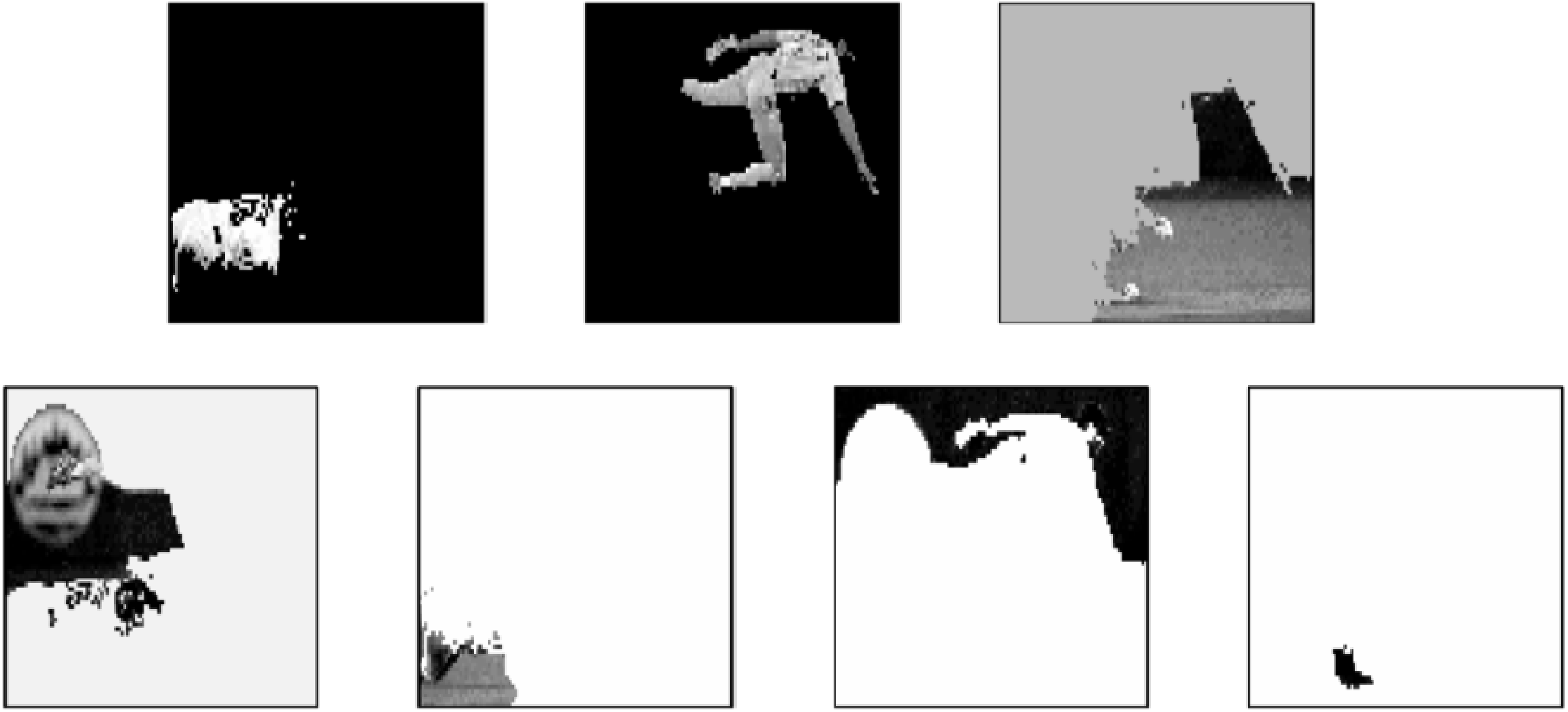
Normalized distance between the *n*th and *m*th pixel





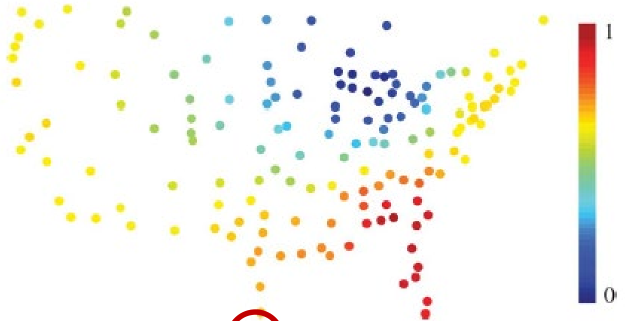
J. Shi and J. Malik, "Normalized cuts and image segmentation*," IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22(8), pp. 888-905, Aug. 2000.

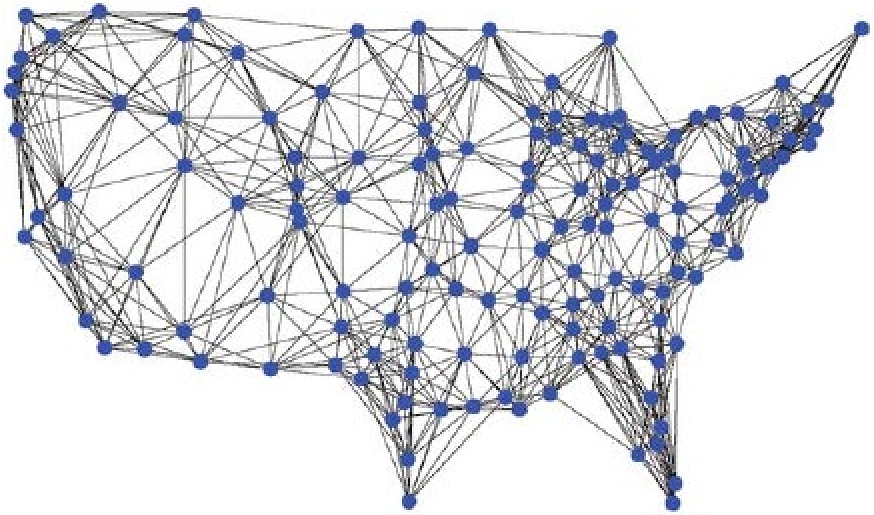
Image Segmentation (Clustering)

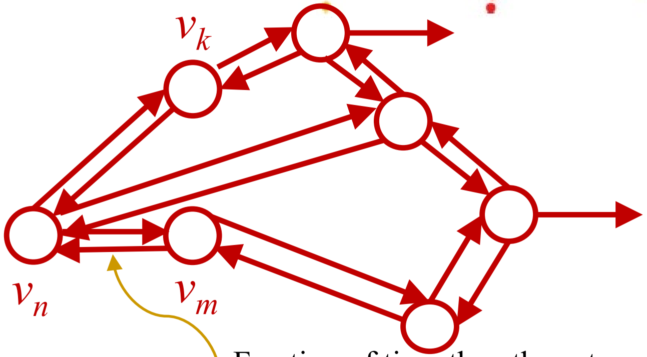


Applications: Customer Behavior Prediction

Determine whether or not existing customers will stop service with a cellular provider





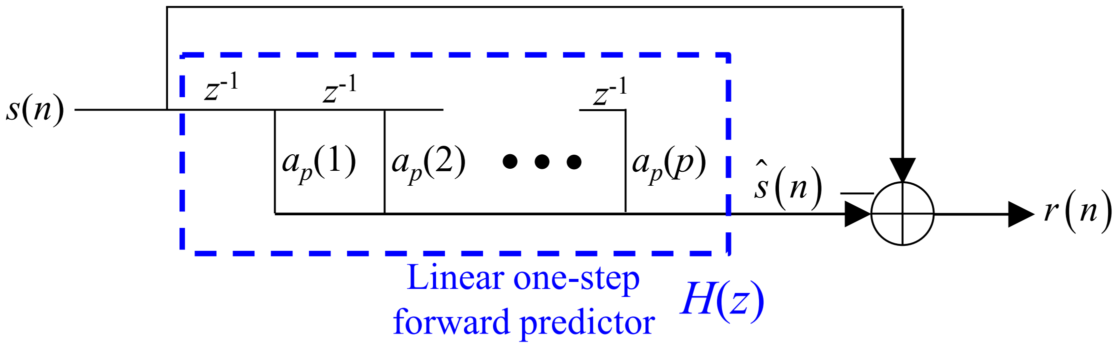


Fraction of time the *n*th customer called and talked to the *m*th customer

$$[\mathbf{A}]_{nm} = \frac{T_{n,m}}{\sum_{k \in \mathcal{N}_n} T_{n,k}}$$

A. Sandryhaila and J.M.F. Moura, ``Discrete signal processing on graphs," *IEEE Trans. on Signal Processing*, vol. 61(7), pp. 1644-1656, Apr. 2013.

Customer Behavior Prediction



$$e[n] = s[n] - \sum_{k=1}^{p} a_k s[n-k], \quad n \geq p$$

$$\Leftrightarrow R(z) = S(z)\left[1 - \sum_{k=1}^{p} a_k z^{-k}\right]$$

$$= S(z)\left[1 - H(z)\right]$$

$$\hat{\mathbf{s}} = h(\mathbf{A})\mathbf{s} \Longrightarrow \mathbf{r} = \left(\mathbf{I}_N - h(\mathbf{A})\right)\mathbf{s}$$

Customer Behavior Prediction

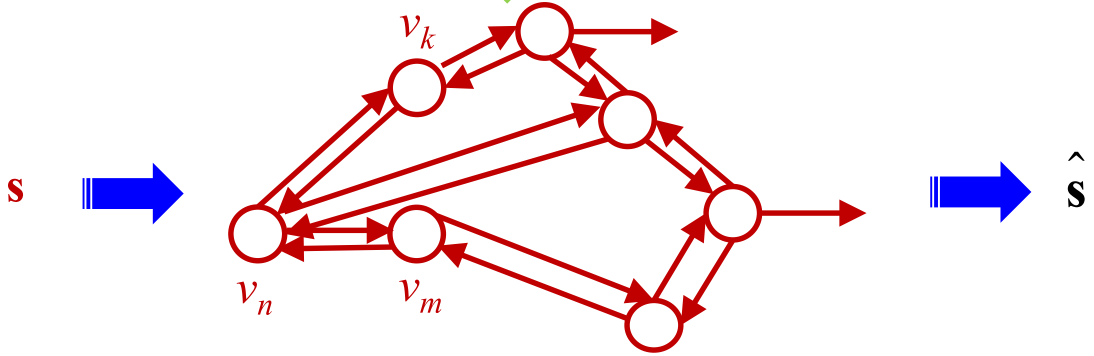
$$\text{Signal: } \begin{cases} [\mathbf{s}]_n = 1, & \text{customer already stopped} \\ [\mathbf{s}]_n = 0, & \text{else} \end{cases}$$

$$h(\mathbf{A}) = h_0\mathbf{I}_N + h_1\mathbf{A} + \cdots + h_{L-1}\mathbf{A}^{L-1} \quad \text{(use an } L\text{-tap filter)}$$

$$\text{Let } h_0 = 1.\ \mathbf{r} = \mathbf{s} - \left(h(\mathbf{A})\mathbf{s}\right) = \left(\mathbf{I}_N - h(\mathbf{A})\right)\mathbf{s}$$

$$\|\mathbf{r}\|_2^2 = \left\|\mathbf{s} - \left(h_1\mathbf{A} + \cdots + h_{L-1}\mathbf{A}^{L-1}\right)\mathbf{s}\right\|_2^2 = \left\|\mathbf{s} - \begin{bmatrix}\mathbf{A}\mathbf{s} & \cdots & \mathbf{A}^{L-1}\mathbf{s}\end{bmatrix}\begin{bmatrix} h_1 \\ \vdots \\ h_{L-1}\end{bmatrix}\right\|_2^2 = \|\mathbf{s} - \mathbf{B}\mathbf{h}\|_2^2$$

$$\mathbf{h}_{opt} = \underset{\mathbf{h}}{\operatorname{argmin}} \|\mathbf{r}\|_2^2 = \left(\mathbf{B}^T\mathbf{B}\right)^{-1}\mathbf{B}^T\mathbf{s}$$



Customer Behavior Prediction

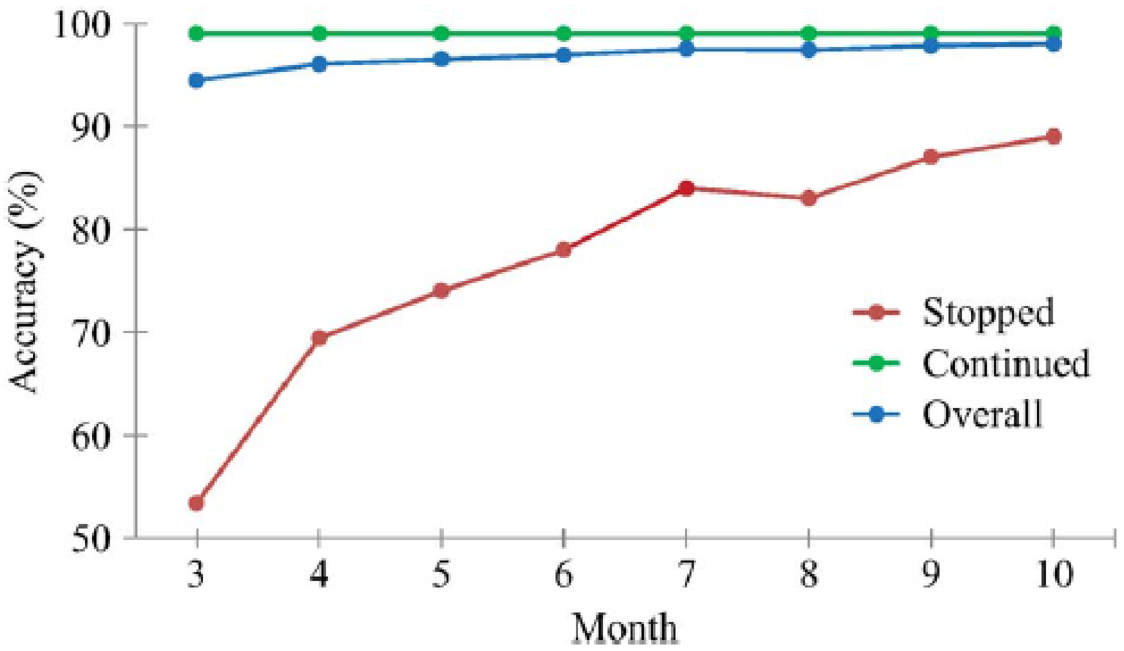


Fig. 7. The accuracy of behavior prediction for customers of a mobile provider. Predictions for customers who stopped using the provider and those who continued are evaluated separately, and then combined into the overall accuracy.

Application: Location-Aware Spatial Received Signal Power Prediction

Signal model

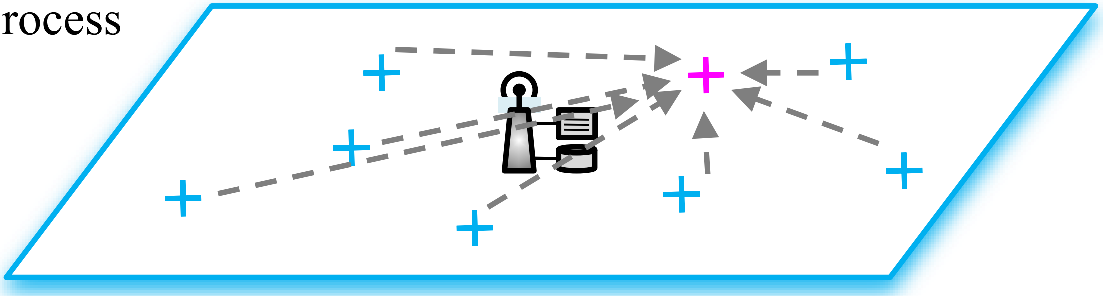
$$\mathbf{y} = \mathbf{p} + \mathbf{w} \in \mathbb{R}^N$$

Log-normal channel model

$$P_{RX}(\mathbf{x}_s, \mathbf{x}_i) = \underbrace{L_0 - 10\eta \log_{10}\left(\|\mathbf{x}_s - \mathbf{x}_i\|_2\right)}_{\text{Path loss}} + \underbrace{\Psi(\mathbf{x}_s, \mathbf{x}_i)}_{\text{Shadowing}}$$

$$\Psi(\mathbf{x}_s, \mathbf{x}_i) \sim \mathcal{N}\left(0, \sigma_\Psi^2\right)$$

p can be modeled as a Gaussian process



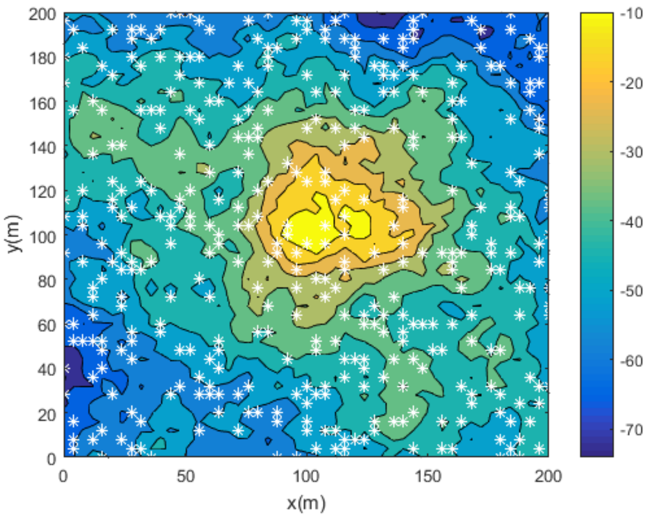
$$C(\mathbf{x}_i, \mathbf{x}_j) = E\left[\Psi(\mathbf{x}_s, \mathbf{x}_i)\Psi(\mathbf{x}_s, \mathbf{x}_j)\right] = \sigma_\Psi^2 \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|_2}{d_c}\right)$$

J. Fink and V. Kumar, “Online methods for radio signal mapping with mobile robots,” *Proc. of the Intl. Conf. on Robotics and Automation*, Anchorage, Alaska, USA, pp. 1940-1945, May 2010.

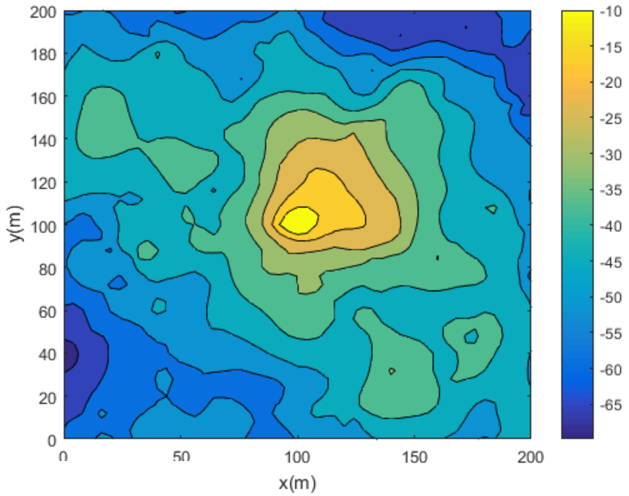
R. Di Taranto *et al.*, “Location-aware communications for 5G networks,” *IEEE Signal Processing Magazine*, vol. 31(6), pp. 102-112, Nov. 2014.

C.C. Fung, C. Liu and R.C. Hung, “Location-aware spatio-temporal received signal power prediction,” *under preparation.*

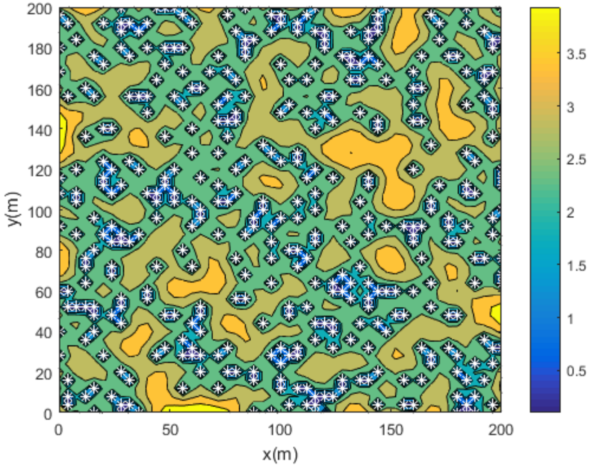
Application: Location-Aware Spatial Received Power Prediction



Actual radio map



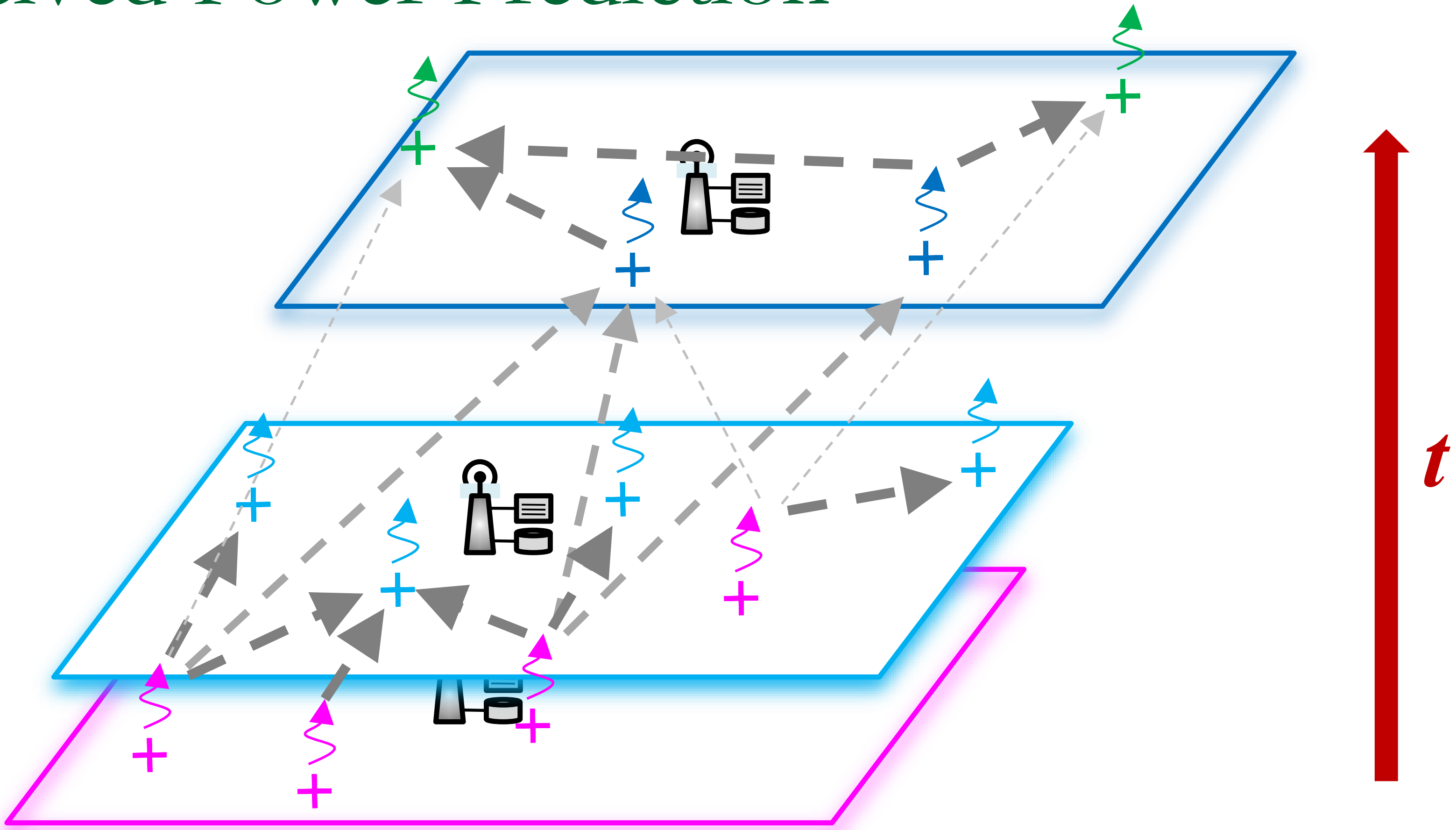
Predicted radio map



Variance of prediction

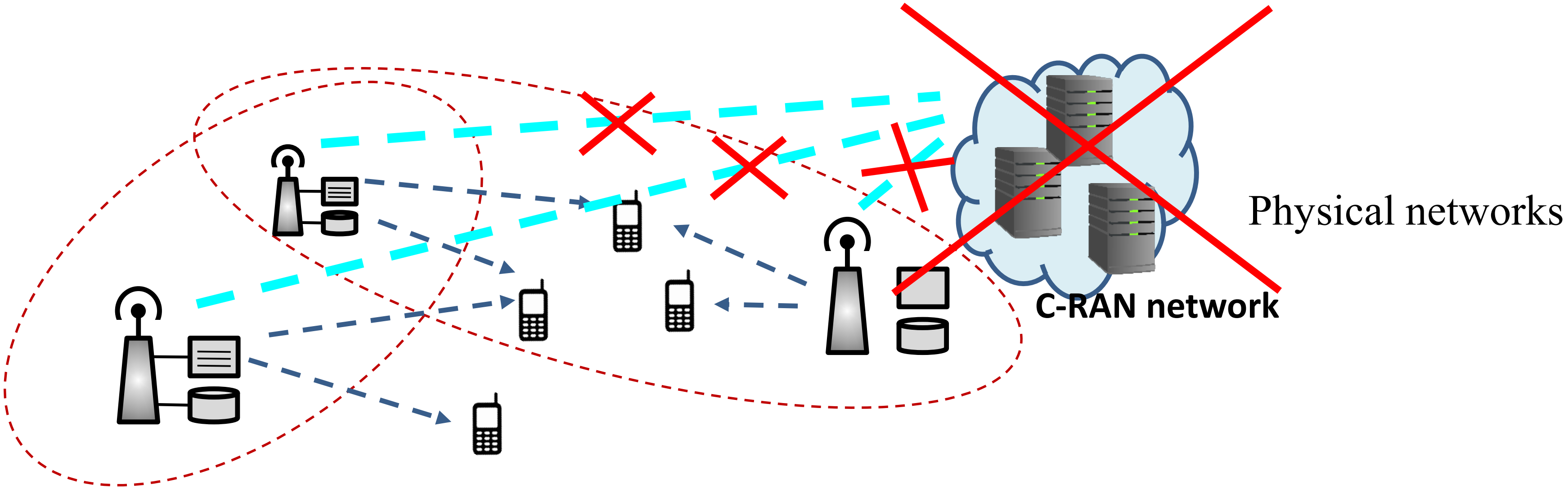
400 training samples (‘*’ sign)
Left: Actual radio map
Right: Predicted radio map
Bottom: Variance of predicted radio map

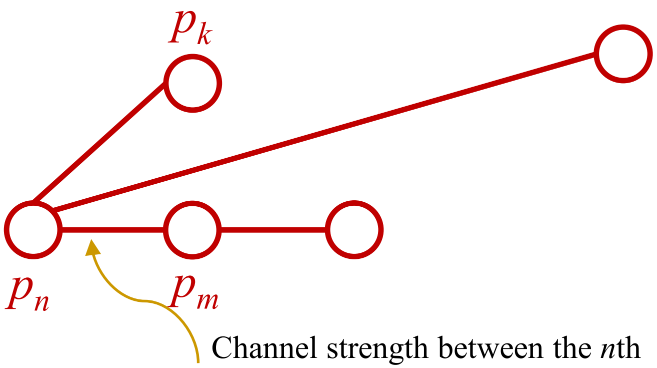
Location-Aware Spatio-Temporal Received Power Prediction



C.C. Fung, C. Liu and R.C. Hung, ``Location-aware spatio-temporal received signal power prediction,'' *under preparation.*

Application: GSP for Distributed Transmitter Design





Channel strength between the *n*th and *m*th TP (transmission point)

GSP for Distributed Transmitter Design

$$\max_{\mathbf{Q}} \sum_{\text{all users}} \sum_{\text{all nodes}} \overbrace{\left(\text{signal power from } j \text{ node} - \text{number of nonzero TPs}\right)}^{f_i^q}$$

$$\text{s.t.} \sum_{\text{all nodes}} \boxed{\text{instantaneous leakage interference from } j \text{ node}}$$

Coupling constraint

g_{ij}^q

$$I_{th} = L_{ij}^{q(n)}$$

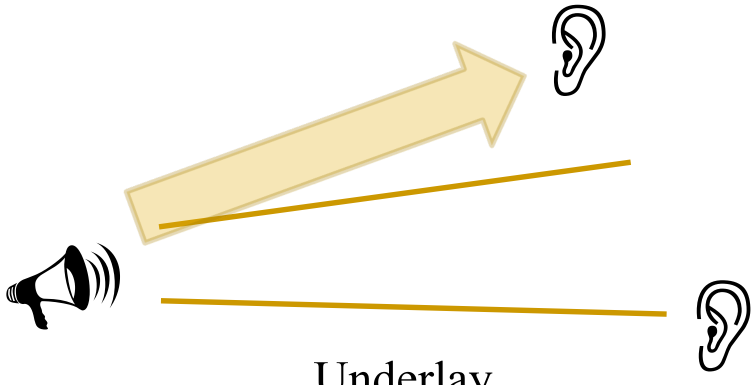
$$\leq \text{ leakage threshold, } \forall \text{users} \neq \text{serving user}$$

$$\sum_{\text{all users}} \text{transmit power at } j \text{ node } \leq \text{ Tx power threshold}$$

Note:

1. If optimizing per node, maximizing each f_i^q term is equivalent to maximizing $\sum_{\text{all nodes}} f_i^q$ if done centrally

2. But the same cannot be said about the coupling constraint



Distributed Optimization Problems

$$\max_{\mathbf{x}_1,\mathbf{x}_2} \sum_{\text{all nodes}} f\left(\mathbf{x}_1,\mathbf{x}_2\right) \quad \text{s.t.} \sum_{\text{all nodes}} g\left(\mathbf{x}_1,\mathbf{x}_2\right) \le 0 \quad \equiv \quad \max_{\mathbf{x}_j} f\left(\mathbf{x}_j\right) \quad \text{s.t.} \sum_{\text{all nodes}} g\left(\mathbf{x}_1,\mathbf{x}_2\right) \le 0$$

How can this problem be resolved?

➡ Dual-consensus algorithm

$$\max_{\mathbf{x}_1,\mathbf{x}_2} \sum_{\text{all nodes}} f\left(\mathbf{x}_1,\mathbf{x}_2\right) \quad \text{s.t.} \sum_{\text{all nodes}} g\left(\mathbf{x}_1,\mathbf{x}_2\right) \le L \quad \ne \quad \max_{\mathbf{x}_j} f\left(\mathbf{x}_j\right) \quad \text{s.t.} \sum_{\text{all nodes}} g\left(\mathbf{x}_1,\mathbf{x}_2\right) - L \le 0$$

➡ "Resource" allocation algorithm

GSP for Distributed Transmitter Design

$\sum_{\text{all nodes}} g_{ij}^{q} \le L_{ij}^{q(n)} \Leftrightarrow \sum_{\text{all nodes}} g_{ij}^{q} - L_{ij}^{q(n)} \le 0$ only when $n = 0$, as we do not know what $L_{ij}^{q(n)}$ for $n > 0$

Let $L_{ij}^{q(0)} = \frac{I_{th}}{\text{total \# of nodes}}$, and redistribute $L_{ij}^{q(n)}$ using proposed "resource allocation" algorithm, for $n > 0$

$$\begin{aligned} \max_{\mathbf{Q}} \quad & \sum_{\text{all users}} \sum_{\text{all nodes}} f_i^q \\ \text{s.t.} \quad & \sum_{\text{all nodes}} g_{ij}^{q} \le L_{ij}^{q}, \ \forall \text{users} \ne \text{serving user} \\ & \sum_{\text{all users}} \text{transmit power at } j \text{ node} \le \text{Tx power threshold} \end{aligned}$$

GSP for Distributed Transmitter Design

Step 1 : Local processing step

$\mathbf{Q}^{q(n+1)} = \underset{\mathbf{Q}^{q(n)}}{\arg\max} \sum_{i} f_i^q + \boldsymbol{\ell}^{q(n)T}\mathbf{g}^{q(n)}$, such that transmit power constraint is satsified

Compute $\tilde{\mathbf{g}}^{q(n+1)} = \mathbf{g}^{q(n)}\left(\mathbf{Q}^{q(n+1)}\right)$ and $L_{ij}^{q(n+1)}$

$$\boldsymbol{\lambda}^{q(n+1)} = \underset{\boldsymbol{\lambda}^q \geq 0}{\arg\min} \sum_{\text{all users}} \boldsymbol{\lambda}^{q(n+1)H}\tilde{\mathbf{g}}^{q(n+1)} - \frac{\left\|\boldsymbol{\lambda}^{q(n)} - \boldsymbol{\ell}^{q(n)}\right\|_2^2}{2c^{(n)}} = P_+\left[\boldsymbol{\ell}^{q(n)} + c^{(n)}\tilde{\mathbf{g}}^{q(n+1)}\right]$$

Step 2 : Communications step

Broadcast $\mathbf{g}^{q(n)}\left(\mathbf{Q}^{q(n+1)}\right) - L_{ij}^{q(n+1)}$, and $\boldsymbol{\lambda}^{q(n+1)}$ to one-hop neigbhors

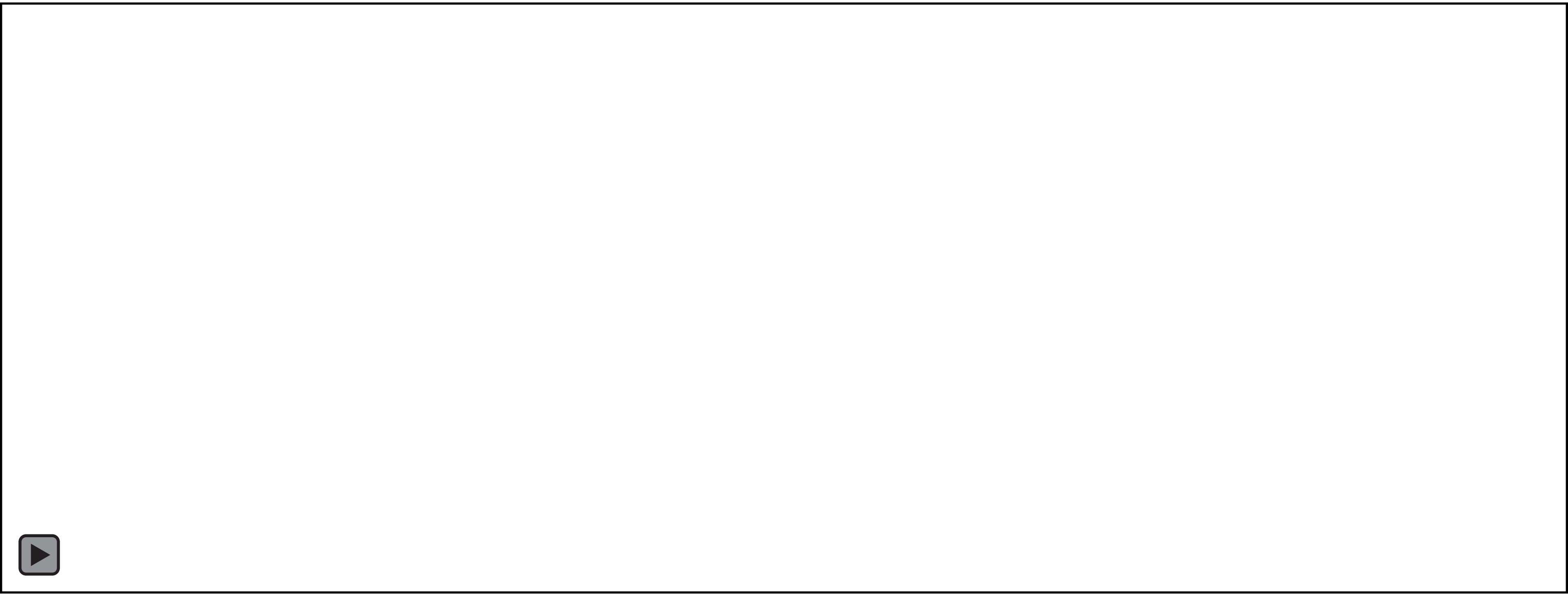
Step 3 : Consensus step

Reallocate interference threshold

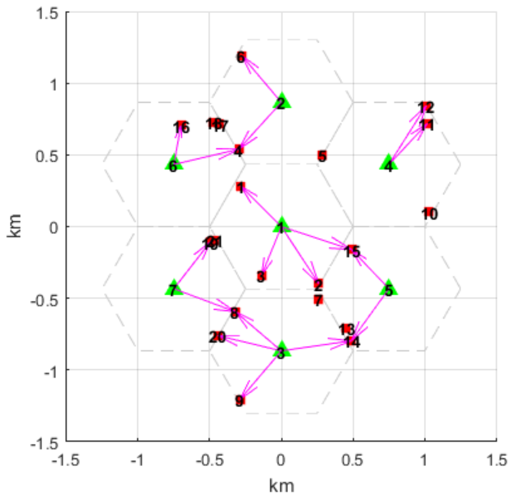
Compute $\boldsymbol{\ell}^{q(n+1)}$ as a weighted sum of all the received $\boldsymbol{\lambda}^{q(n+1)}$'s from one-hop neighbors

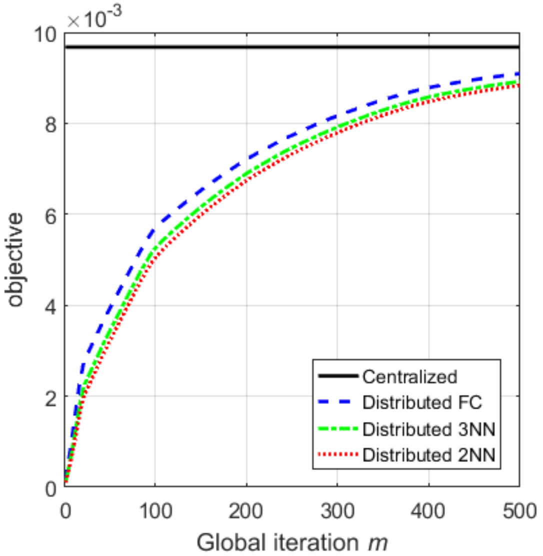
M. Servetnyk and **C.C. Fung**, “Distributed joint transmitter design and selection for heterogeneous networks using consensus-based dual decomposition,” *submitted to the IEEE Trans. on Signal Processing*, 2018.
M. Servetnyk and **C.C. Fung**, “Distributed joint transmitter design and selection using augmented ADMM,” *submitted to the IEEE Intl. Conf. on Acoustics, Speech and Signal Processing*, 2018.

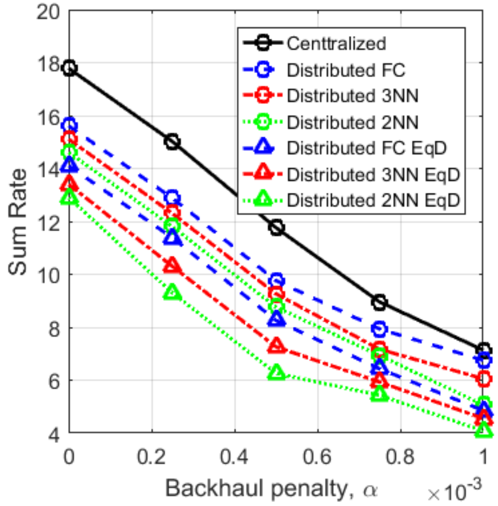
Distributed JT-CoMP



GSP: Distributed JT-CoMP







Conclusions

- Graph model is versatile in representing data in a structure way
  - Can represent signal, information and physical network where data lie on an irregular structure
    - Physical network allows for distributed processing
  - Unfortunately, it is not intuitive in representing causality
- Many classical signal processing approaches have graph counterpart
- Graph models can solve many issues regarding Big Data
- Open questions:
  - How do we popular the graph (graph learning)
    - No initial graph exists
    - Or modification of known graph is necessary
    - Which topology is used?
  - How do get the best bang out of graphs for each application

What skills are required/learned to be successful?

- Good in mathematics and programming
  - Optimization, graph theory (graph signal processing), statistics, Matlab+Python/Julia(?)
- Willingness and courage to explore and learn new (cross-disciplinary) subjects
- Ingenuity
- Be vocal

c.fung@ieee.org
http://cwww.ee.nctu.edu.tw/~cfung
or Google "Carrson Fung"