Optimal Network Decomposition for Next-Generation Mobile Communication Systems

Lin Dai

Department of Electronic Engineering
City University of Hong Kong

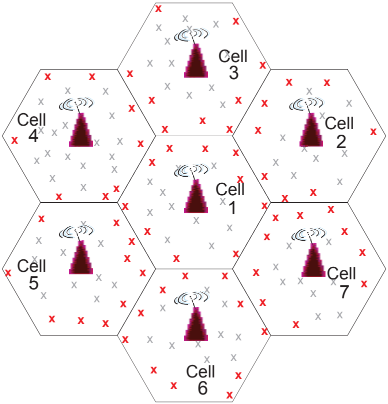
Collaborative Work with Dr. Bo Bai from Huawei Technologies Co., Ltd.

April 2019

Mobile Communication Systems: From 1G to 5G

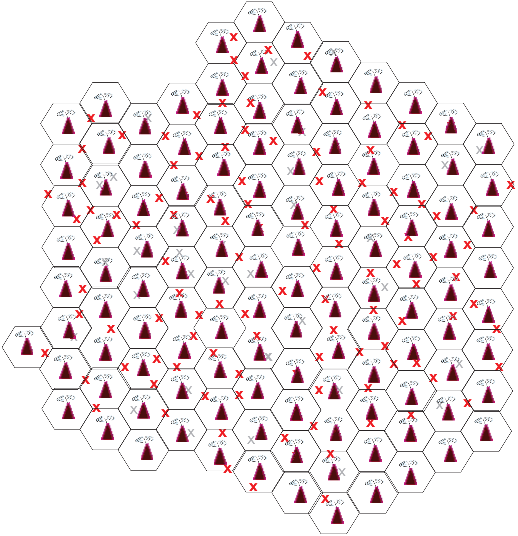
- More Bandwidth
- More Antennas
- More Base Stations (BSs)
- More Sophisticated Coding/Decoding Schemes
- FDMA $\rightarrow$ TDMA $\rightarrow$ CDMA $\rightarrow$ OFDMA

The Network Structure Has Remained Unchanged So Far...



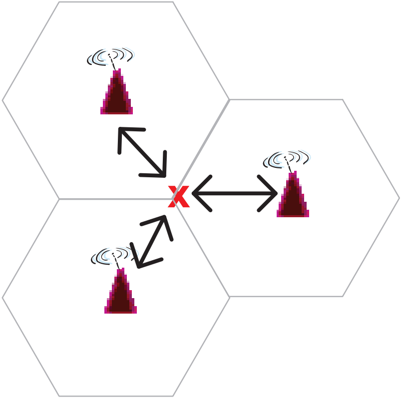
- All adopt the cellular structure, despite drastic changes in the physical and multiple access layers.
- A large area is divided into a number of cells where a base station (BS) is placed at the center of each cell and serves users who fall into its coverage.
- Cell-edge Problem: Users at the cell-edge area suffer from strong interference from neighboring BSs.

The Network Structure Has Remained Unchanged So Far...



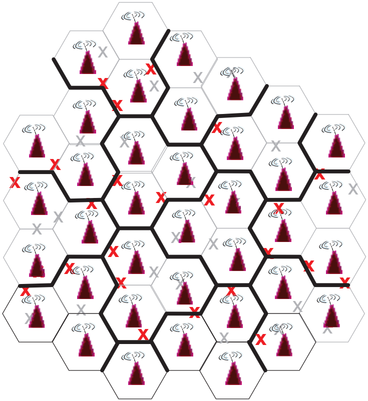
- Explosive growth in network size and traffic demand
  $\rightarrow$
  Increasing density of BSs
  $\rightarrow$
  Shrinking cell size
- More users fall into the cell-edge area as the size of each cell shrinks!

Solution I: BS Coordination



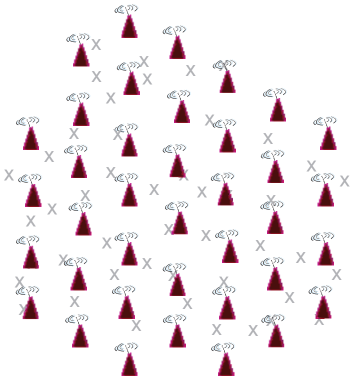
- Multiple BSs are clustered together, and jointly serve the cell-edge users.

Solution I: BS Coordination



- Multiple BSs are clustered together, and jointly serve the cell-edge users.
- Clusters are formed from the BS side $\rightarrow$ Cluster-edge users

Solution II: Cell-Free



- Distributed Antenna System (DAS) / Cloud Radio Access Network (C-RAN) / Cell-Free Network:

  BSs (remote radio heads) are connected to a central controller and jointly serve all the users.

- Unscalable: Computational complexity ↑ as the densities of BSs and users ↑

Network Decomposition

- The fundamental idea of network decomposition is to break a large-scale network into smaller parts such that subnetworks can operate in parallel, each with lower dimensionality.
- It was originally proposed for electric power systems, and has been widely applied to many engineering disciplines such as circuit design and network reliability.
- For a large complex network, global control and optimization is typically infeasible. It is highly desirable to decompose the network into many subnetworks, and keep the interconnection among subnetworks at a low level.

Network Decomposition

- Both the cellular structure and BS coordination are examples of network decomposition, but only based on the BS topology.
- BS-centric decomposition schemes lead to increasingly significant interference among subnetworks as the densities of BSs and users grow.
- Efficient network decompositions schemes must be developed to replace the current cellular structure!

Optimal Network Decomposition for Infrastructure-based Wireless Networks

- Intuitively, the network decomposition should be performed based on the geographical information of both BSs and users.
- Proposed network decomposition theory from a graph-theoretic point of view:

  Based on a novel bipartite graph representation of an infrastructure-based wireless network, it is shown that in general the optimal network decomposition can be formulated as a bipartite graph partitioning problem.

Optimal Network Decomposition for Infrastructure-based Wireless Networks

- A Graph-Theoretic Framework
- Binary Search based Spectral Relaxation (BSSR) Algorithm
- Insights to 5G Network Design

A Graph-Theoretic Framework

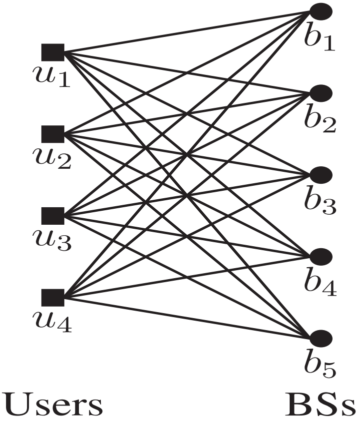
Graph Theory in Wireless Communications

- Most focused on ad-hoc (infrastructure-less) networks for scheduling, routing and connectivity problems.

- For large-scale infrastructure-based wireless networks, optimization techniques developed in graph coloring and matching were applied to solve various resource allocation problems.

  In most of the studies, an interference graph is established where the vertices are either BSs or mobile users, and the edge weight depends on the interference level between two vertices whose definition varies with system assumptions.

A Bipartite Graph Representation



- For an infrastructure-based wireless network, there are two types of nodes, i.e., BSs and users, and a connection only exists between two nodes of different types.
- It can be modeled as a weighted bipartite graph $\mathcal{G}(\mathcal{U} \cup \mathcal{B}, \mathcal{E})$, where $\mathcal{U} = \{u_1, u_2, \cdots, u_K\}$ and $\mathcal{B} = \{b_1, b_2, \cdots, b_L\}$ denote the set of users and the set of BSs, respectively, and the edge set $\mathcal{E} = \{(u_k, b_l), u_k \in \mathcal{U}, b_l \in \mathcal{B}\}$.

The optimal network decomposition of an infrastructure-based wireless network can be formulated as a bipartite graph partitioning problem.

Review of Graph Partitioning Algorithms

According to whether the number of subgraphs is given, the existing algorithms can be categorized into two classes:

- Class-One: Divide the graph into a given number of subgraphs with equal or similar sizes while minimizing the cut of subgraphs.
  – Examples: k-way partitioning, k-means clustering, spectral clustering, ...
- Class-Two: Group vertices based on their similarity.
  – Examples: Hierarchical clustering, affinity propagation, ...

Graph Partitioning for Infrastructure-based Wireless Networks

- For infrastructure-based wireless networks, the number of decomposed subnetworks determines a tradeoff between complexity and network performance: The larger number of decomposed subnetworks, the lower system complexity for each subnetwork, but the higher interconnection, i.e., interference, among subnetworks.

  – We are interested in the case that the number of decomposed subnetworks is maximized given that the interference among subnetworks is bounded, which can be categorized as a class-two graph partitioning problem.

- For infrastructure-based wireless networks, similarity between two BSs or mobile users has little physical meaning.

  – We adopt a double-loop approach to convert the class-two graph partitioning problem into class-one, and then solve it based on spectral clustering.

Notations

Consider a subset $\mathcal{C}$ of the vertex set $\mathcal{V} = \mathcal{U} \cup \mathcal{B}$, and define $\mathcal{U}_{\mathcal{C}} = \mathcal{U} \cap \mathcal{C}$, $\overline{\mathcal{U}}_{\mathcal{C}} = \mathcal{U} \setminus \mathcal{U}_{\mathcal{C}}$, $\mathcal{B}_{\mathcal{C}} = \mathcal{B} \cap \mathcal{C}$, and $\overline{\mathcal{B}}_{\mathcal{C}} = \mathcal{B} \setminus \mathcal{B}_{\mathcal{C}}$.

- Define the cut function $\mathrm{cut}(\mathcal{C})$ as $\mathrm{cut}(\mathcal{C}) = \sum_{u_k \in \mathcal{U}_{\mathcal{C}}, b_l \in \overline{\mathcal{B}}_{\mathcal{C}}} w_{kl} + \sum_{b_l \in \mathcal{B}_{\mathcal{C}}, u_k \in \overline{\mathcal{U}}_{\mathcal{C}}} w_{kl}$.
- Define the volume function $\mathrm{vol}(\mathcal{C})$ as $\mathrm{vol}(\mathcal{C}) = \sum_{u_k \in \mathcal{U}_{\mathcal{C}}, b_l \in \mathcal{B}} w_{kl} + \sum_{b_l \in \mathcal{B}_{\mathcal{C}}, u_k \in \mathcal{U}} w_{kl}$.
- $\mathrm{cut}(\mathcal{C})$ and $\mathrm{vol}(\mathcal{C}) - \mathrm{cut}(\mathcal{C})$ represent the dissimilarity of subgraphs and similarity within a subgraph, respectively. For wireless networks, they can be interpreted as the interference between nodes inside subnetwork $\mathcal{C}$ and outside subnetwork $\mathcal{C}$, and the signal power of nodes inside subnetwork $\mathcal{C}$, respectively.

Problem Formulation

The optimal network decomposition can then be formulated as the following *Maximum-Number* (*Max-Num*) partitioning problem:

$$\text{(P1)} \qquad \max_{\Pi_M(\mathcal{G})} \quad M \qquad \text{s.t.} \quad \sum_{j=1}^{M} \frac{\mathrm{cut}(\mathcal{C}_j)}{\mathrm{vol}(\mathcal{C}_j) - \mathrm{cut}(\mathcal{C}_j)} \leq \delta,$$

for some $\delta \geq 0$, where $\Pi_M(\mathcal{G})$ denotes a partition of the weighted bipartite graph $\mathcal{G}$.

In the constraint, we consider the ratio $\frac{\mathrm{cut}(\mathcal{C}_j)}{\mathrm{vol}(\mathcal{C}_j)-\mathrm{cut}(\mathcal{C}_j)}$ instead of the $\mathrm{cut}(\mathcal{C}_j)$ because:

1. From the graph partitioning perspective, it leads to a more balanced partitioning result;

2. From the application side, the network performance of a wireless network is determined by the ratio of the interference to the signal power, rather than the absolute interference level.

Binary Search based Spectral Relaxation (BSSR) Algorithm

Solve P1 in Two Loops

A double-loop approach for solving the max-num partitioning problem P1:

- *Inner-Loop*: For a given number of subgraphs M, find the optimal partition $\Pi_M^*(\mathcal{G}) = \{\mathcal{C}_1^*, \mathcal{C}_2^*, \cdots, \mathcal{C}_M^*\}$ to minimize the sum of the ratio $\frac{\mathrm{cut}(\mathcal{C}_j)}{\mathrm{vol}(\mathcal{C}_j) - \mathrm{cut}(\mathcal{C}_j)}$:

$$\{\mathcal{C}_1^*, \mathcal{C}_2^*, \cdots, \mathcal{C}_M^*\} = \operatorname{argmin} \sum_{j=1}^{M} \frac{\mathrm{cut}(\mathcal{C}_j)}{\mathrm{vol}(\mathcal{C}_j) - \mathrm{cut}(\mathcal{C}_j)}; \tag{1}$$

- *Outer-Loop*: Find the maximum number of subgraphs:

$$M^* = \max \left\{ M : \sum_{j=1}^{M} \frac{\mathrm{cut}(\mathcal{C}_j^*)}{\mathrm{vol}(\mathcal{C}_j^*) - \mathrm{cut}(\mathcal{C}_j^*)} \leq \delta \right\}. \tag{2}$$

Preliminaries: Outer-Loop

Theorem 1 shows that $\sum_{j=1}^{M} \frac{\mathrm{cut}(\mathcal{C}_j^*)}{\mathrm{vol}(\mathcal{C}_j^*)-\mathrm{cut}(\mathcal{C}_j^*)}$ is a monotone increasing function of M, which enables a binary search of M^*.

Theorem

$$\min_{\Pi_M(\mathcal{G})} \sum_{j=1}^{M} \frac{\mathrm{cut}(\mathcal{C}_j)}{\mathrm{vol}(\mathcal{C}_j) - \mathrm{cut}(\mathcal{C}_j)} \leq \min_{\Pi_{M+1}(\mathcal{G})} \sum_{j=1}^{M+1} \frac{\mathrm{cut}(\mathcal{C}_j)}{\mathrm{vol}(\mathcal{C}_j) - \mathrm{cut}(\mathcal{C}_j)}.$$

More notations:

- The adjacency matrix $\mathbf{A}$ of graph $\mathcal{G}$ is
$$\mathbf{A} = \begin{bmatrix} \mathbf{0}_{K \times K} & \mathbf{W} \\ \mathbf{W}^T & \mathbf{0}_{L \times L} \end{bmatrix}, \tag{3}$$
where $\mathbf{W}$ denotes the weight matrix of edge set $\mathcal{E}$.
- $\mathbf{D} = \mathrm{diag}\left(d_1, d_2, \ldots, d_{K+L}\right)$ is a diagonal matrix whose i-th diagonal entry is given by
$$d_i = \sum_{v_j \in \mathcal{V}} a_{ij}, \tag{4}$$
i.e., the weighted degree of vertex $v_j \in \mathcal{V}$.

More notations:

- Define a decision matrix $\mathbf{X}$ whose entry in the i-th row and j-th column x_{ij} is given by

$$x_{ij} = \begin{cases} \dfrac{1}{\sqrt{\mathrm{vol}\,(\mathcal{C}_j)}}, & \text{if} \quad v_i \in \mathcal{C}_j; \\ 0, & \text{otherwise}, \end{cases} \tag{5}$$

  where vertex $v_i \in \mathcal{V}$, $i = 1, \ldots, K + L$, and $j = 1, \ldots, M$.
- Denote the j-th column of $\mathbf{X}$ as $\mathbf{x}_j^c$, $j = 1, \ldots, M$, and the i-th row of $\mathbf{X}$ as $\mathbf{x}_i^r$, $i = 1, \ldots, K + L$.

The optimization problem (1) is equivalent to

$$(\mathrm{P2}) \quad \begin{aligned} \min_{\mathbf{X}} \quad & \sum_{j=1}^{M} \frac{1}{(\mathbf{x}_j^c)^T \mathbf{A} \mathbf{x}_j^c} - M \\ \text{s.t.} \quad & (\mathbf{x}_j^c)^T \mathbf{D} \mathbf{x}_j^c = 1, \qquad j = 1, \ldots, M, \\ & \|\mathbf{x}_i^r\|_0 = 1, \qquad i = 1, \ldots, K+L, \end{aligned} \tag{6}$$

which is NP-hard due to the last set of constraints.
If we relax $\mathrm{P2}$ to

$$(\mathrm{P2}') \quad \begin{aligned} \min_{\mathbf{X}} \quad & \sum_{j=1}^{M} \frac{1}{(\mathbf{x}_j^c)^T \mathbf{A} \mathbf{x}_j^c} - M \\ \text{s.t.} \quad & (\mathbf{x}_j^c)^T \mathbf{D} \mathbf{x}_j^c = 1, \qquad j = 1, \ldots, M, \end{aligned} \tag{7}$$

then the optimal solution $\tilde{\mathbf{X}}^*$ of $\mathrm{P2}'$ can be obtained as $\tilde{\mathbf{X}}^* = [\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_M]$, i.e., the eigenvectors corresponding to the smallest M eigenvalues of the random-walk graph Laplacian matrix $\mathbf{L}_{\mathrm{rw}} = \mathbf{D}^{-1}(\mathbf{D} - \mathbf{A})$.

Preliminaries: Inner-Loop

Theorem 2 shows that the optimal solution $\tilde{\mathbf{X}}^*$ of $\mathrm{P2}'$ is equal to the optimal solution $\mathbf{X}^*$ of $\mathrm{P2}$ if the multiplicity of zero-eigenvalue N of the random-walk graph Laplacian matrix $\mathbf{L}_{\mathrm{rw}}$ is no smaller than the number of subgraphs M.

Theorem

If $M \leq N$, then $\|\tilde{\mathbf{x}}_i^{r*}\|_0 = 1$, $i = 1, \ldots, K + L$.

BSSR Algorithm

Inner-loop:

- Obtain the optimal solution $\tilde{\mathbf{X}}^*$ of $\mathrm{P2}'$: $\tilde{\mathbf{X}}^* = [\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_M]$;
- Approximate the optimal solution $\mathbf{X}^*$ of $\mathrm{P2}$ based on $\tilde{\mathbf{X}}^*$;
- Obtain the optimal partition $\Pi_M^*(\mathcal{G}) = \{\mathcal{C}_1^*, \mathcal{C}_2^*, \cdots, \mathcal{C}_M^*\}$ based on $\mathbf{X}^*$.

Outer-loop: Binary search.

Algorithm 1 BSSR Algorithm

```
Input the weighted adjacency matrix A.
Compute the eigenvectors p_1, p_2, ..., p_{K+L} of L_rw.
low ← 1, high ← min{K, L}
repeat
    M ← ⌊(low+high)/2⌋
    X̃* ← [p_1, p_2, ..., p_M]
    Compute X* iteratively based on (17)-(18).
    Compute Π*_M(G) = {C*_1, C*_2, ..., C*_M} based on (11).
    if Σ_{j=1}^{M} cut(C*_j) / (vol(C*_j) − cut(C*_j)) ≤ δ then
        low ← M + 1
    else
        high ← M − 1
    end if
until low > high
M* ← M, Π*_{M*}(G) ← Π*_M(G)
```

Optimality of BSSR

- $\Pi^*_{M^*}(\mathcal{G})$ obtained by the proposed BSSR algorithm is the optimal partition for the max-num problem $\mathrm{P1}$ if and only if $\tilde{\mathbf{X}}^*$ is equal to the optimal solution $\mathbf{X}^*$ of $\mathrm{P2}$.
- In general, there is no guarantee of the optimality due to the spectral relaxation introduced in the inner-loop.
- The proposed BSSR algorithm is optimal if the sum ratio threshold $\delta = 0$:

  When $\delta = 0$, $\tilde{\mathbf{X}}^*$ is equal to the optimal solution $\mathbf{X}^*$ of $\mathrm{P2}$, and the maximum number of subgraphs M^* is given by the multiplicity of zero-eigenvalue of the random-walk graph Laplacian matrix.

Complexity of BSSR

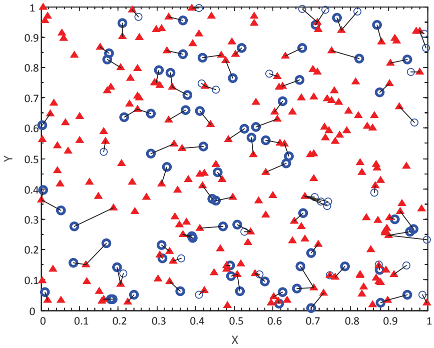
Table : Complexity Comparison of BSSR with $\delta > 0$ and $\delta = 0$

	Eigenvalue Decomposition	Number of Iterations
$\delta > 0$	$O((K+L)^3)$	$O(\log \min(K, L))$
$\delta = 0$	$O(KL)$	1

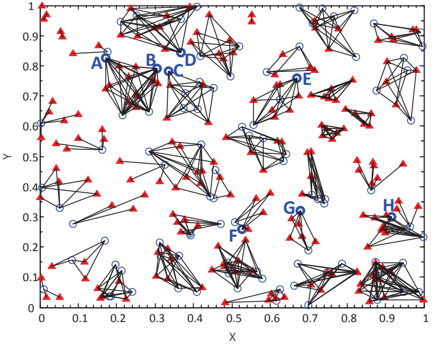
Insights to 5G Network Design

Performance Benchmarks

- Cellular Structure



- BS Clustering (k-means++)



The decomposition is based on a randomly generated network topology. $L = 200$ BSs and $K = 100$ users are represented by triangles and circles, respectively.

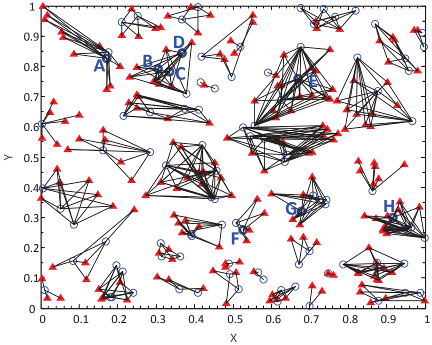
Channel-Gain-based BSSR

- Set the weight w_{kl} of edge (u_k, b_l) as the channel gain $g_{kl} = d_{kl}^{-\alpha}$ between user $u_k \in \mathcal{U}$ and BS $b_l \in \mathcal{B}$.
- The cut ratio of each subnetwork $\mathcal{C}_j$ can be written as

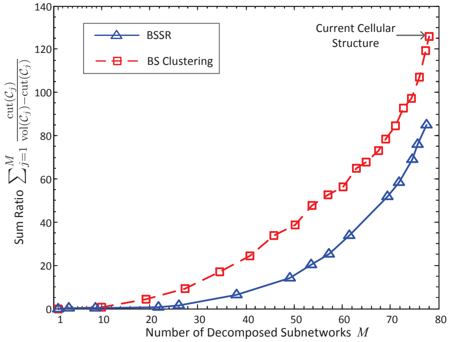
$$\frac{\mathrm{cut}(\mathcal{C}_j)}{\mathrm{vol}(\mathcal{C}_j) - \mathrm{cut}(\mathcal{C}_j)} = \frac{1}{2}\left(\frac{\sum_{u_k \in \mathcal{U}_{\mathcal{C}}, b_l \in \overline{\mathcal{B}}_{\mathcal{C}}} d_{kl}^{-\alpha}}{\sum_{u_k \in \mathcal{U}_{\mathcal{C}}, b_l \in \mathcal{B}_{\mathcal{C}}} d_{kl}^{-\alpha}} + \frac{\sum_{b_l \in \mathcal{B}_{\mathcal{C}}, u_k \in \overline{\mathcal{U}}_{\mathcal{C}}} d_{kl}^{-\alpha}}{\sum_{b_l \in \mathcal{B}_{\mathcal{C}}, u_k \in \mathcal{U}_{\mathcal{C}}} d_{kl}^{-\alpha}}\right), \quad (8)$$

$j = 1, \ldots, M$. With equal transmission power of each user/BS, the first and second items of the right-hand side of (8) represent the ratio of the total interference to the signal power of subnetwork $\mathcal{C}_j$ in the downlink and uplink, respectively.

- Decomposition Result of BSSR

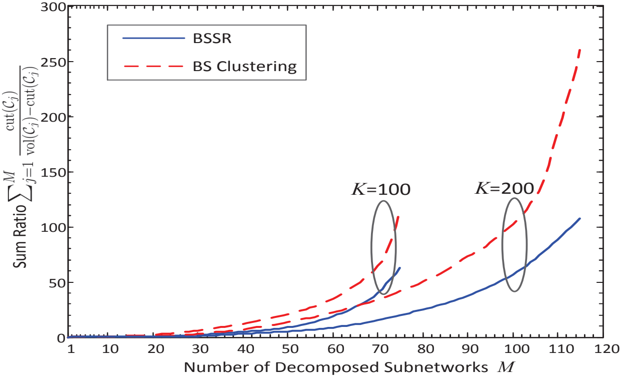


- Sum Ratio Comparison



Channel-Gain-based BSSR

- Averaged over 50 randomly generated network topologies.
- Gain over BS clustering ↑ as the number of subnetworks M or the number of users K ↑



Connection-based BSSR

- To implement the channel-gain-based BSSR, all the $K \times L$ channel gains need to be measured and fed back to the central controller, which leads to heavy signaling overhead for a large-scale system.

- By approximating these insignificant channel gains as zero, the weight matrix $\mathbf{W}$ becomes sparse, based on which the network decomposition can be greatly simplified.

  Let $\tilde{\mathbf{W}}$ denote the 0-1 discretized version of the weight matrix $\mathbf{W}$, which can be regarded as a connection matrix which indicates the connection mapping between users and BSs, i.e., $\tilde{w}_{kl} = 1$ if user u_k is served by BS b_l and $\tilde{w}_{kl} = 0$ otherwise.

  As $\tilde{\mathbf{W}}$ has multiple zero entries, the network decomposition can be performed with the threshold $\delta = 0$, in which case the proposed BSSR algorithm can be greatly simplified and the computational complexity is determined by the number of non-zero entries of $\tilde{\mathbf{W}}$.

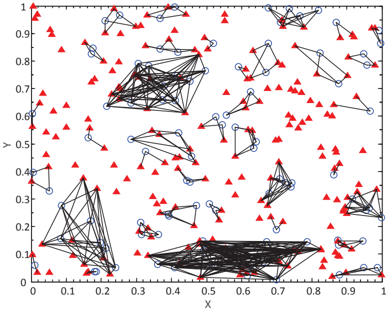
Connection-based BSSR

- There are many ways to discretize the channel-gain matrix $\mathbf{W}$ into the connection matrix $\tilde{\mathbf{W}}$.

- Virtual-cell-based: $\tilde{w}_{kl} = 1$ if BS $b_l \in \mathcal{V}_k$ and $\tilde{w}_{kl} = 0$ otherwise, where $\mathcal{V}_k$ is the virtual cell of user $k \in \mathcal{K}$.

  $\mathcal{V}_k = \{b_l \in \mathcal{B} : w_{kl} \in \{w_{k,1:L}, \cdots, w_{k,V:L}\}\}$, with $w_{k,1:L} \geq w_{k,2:L} \geq \cdots \geq w_{k,L:L}$ denoting the ordered statistics of the channel gains from L BSs to user k.

- Decomposition Result with $V = 2$



- Number of Subnetworks and Sum Ratio versus Virtual Cell Size

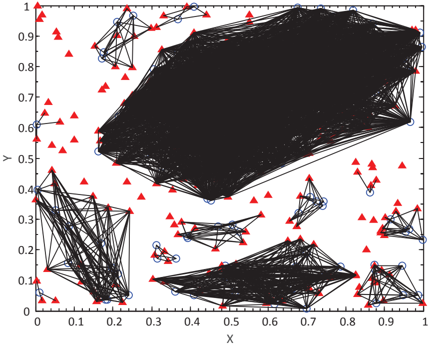
Virtual Cell Size V	Number of Decomposed Subnetworks	Sum Ratio
2	37	8.115
3	13	0.3586
4	6	0.1408
5	4	0.1185
6	3	0.06867
7	1	0

Connection-based BSSR versus Channel-Gain-based BSSR

- Sum ratio: For given number of decomposed subnetworks M, the connection-based BSSR always has a higher sum ratio than the channel-gain-based BSSR, but the gap diminishes as M decreases.
- Complexity: $O(KV)$ for Connection-based BSSR and $O((K+L)^3)$ for Channel-Gain-based BSSR.
- Similar to the sum ratio threshold δ in the channel-gain-based BSSR, here the virtual cell size V determines a tradeoff between decomposability and interference as well, though at a coarser granularity level.

More about Connection-based BSSR

- Connection-based BSSR may lead to unbalanced cluster sizes when the virtual cell size V is large.
- To reduce the imbalance of cluster sizes, additional time/frequency resources need to be allocated to some BSs/users to break the giant component into smaller ones.



Decomposition Result with $V = 3$

Summary

- A first attempt to study the optimal decomposition of large-scale infrastructure-based wireless networks.
- By modeling the network as a bipartite graph, the optimal decomposition was formulated as a maximum-number graph partitioning problem, and solved in two loops.
- The BSSR algorithm was proposed based on the monotone property of the constraint function in the outer-loop and spectral relaxation in the inner-loop, and proven to be optimal when the sum ratio threshold is zero.

Summary

- To demonstrate the practical significance of optimal network decomposition, the performance of the proposed BSSR algorithm was examined by setting the edge weight as the channel gain, and shown to outperform the current cellular structure and BS clustering with an increasing gap as the number of users grows.
- Substantial gains achieved by the proposed BSSR over the current cellular structure and BS clustering in various scenarios corroborate that the optimal network decomposition of next-generation cellular networks should be performed based on a bipartite graph that includes the geographical information of both BSs and users.

Summary

- To reduce the signalling overhead and computational complexity, it was further proposed to implement the BSSR algorithm based on a binary connection matrix.
- The connection-based BSSR was evaluated in a virtual-cell-based next-generation cellular system, and found to achieve close performance to the channel-gain-based BSSR with much lower complexity, but at the cost of unbalanced cluster sizes.

On-going and Future Work

- More refined constraints
- Characterization of network decomposability
- Low-complexity/distributed decomposition schemes
- Disruptive changes to precoding and access design

L. Dai and B. Bai, "Optimal Decomposition for Large-Scale Infrastructure-Based Wireless Networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 4956–4969, Aug. 2017.

The End

Thank You!