I. Motivation

- Next generation content services: Higher data rates
- Wireless bottleneck
- Exploit memory at the wireless edge: Caching

- Goal: Lower the network load for a given cache memory.
- Approach: Exploit correlation across library files.
  - Episodes of a TV show, similar sports events

Correlation across files → Reduction in network load.

II. Caching, Multicast and Coding

- Caching: Prefetch popular content → reduce network load
  - Caching Phase: network resources abundant, demands not known
  - Delivery Phase: network resources limited, demands revealed

- Coded Multicast:
  - Simultaneously serve distinct requests leveraging caches

- Correlation: Potential load reduction

- Literature: Correlated Library
  - Two receivers and one cache: [Timo, Bidokhti, Wigger, Geiger, arxiv, 2016]
  - No coding opportunity
  - Two files and two users: [Hassanzadeh, Tulino, Llorca, Erkip, IST, 2016]
  - No optimality

III. Two-Step Scheme

- Separate compression and caching/delivery
  - Correlation-Aware Cache-Aided Scheme

  - Step 1: Library Compression
    - Gray-Wyner Source Coding:
      - File X1
      - X2
      - X3
      - Gray-Wyner Encoder

    - Decoders 1, 2, 3

    - Multicast Codeword

  - Step 2: Multiple-Request Caching Scheme

  - Gray-Wyner Source Coding

IV. Two Request Scheme

- Three files and two users: cache size \( \frac{1}{2} \rho' \)
  - Uncoded prefetching: load \( \frac{7}{3} \rho' \)
  - Coded prefetching: load \( 2 \rho' \)

V. Performance and Discussions

- A lower bound on the optimal peak rate-memory trade-off.
- The gap between the achievable rate and the lower bound.

- Special Source: Independent components
  - \( X_1 = (U, V_x, V_y) \)
  - \( X_2 = (U, V_x, V_y) \)
  - \( X_3 = (U, V_x, V_y) \)

- Caching Strategy:
  - Low cache capacity: prioritize the middle library.
  - High cache capacity: prioritize the first two libraries.