Available Online: 3 November 2026 DOI: 10.54254/3029-0880/2026.38602

Associations between intrinsic matroid structure and minimum finite field requirements

Guanxin Xiong

School of Glasgow, University of Electronic Science and Technology of China, Chengdu, China

xtj6751@outlook.com

Abstract. Finite field representability connects matroid theory with linear network coding, yet variation in minimum field size across matroids lacks a structural explanation. This study examines the association between intrinsic structure and minimum field size. It introduces a duality invariant Matroid Structural Signature (V5-MSS) and a Finite Field Requirement Correlation Index (FRCI). The signature combines dual rank imbalance, intrinsic dependence, circuit and cocircuit coupling, and local geometric capacity. FRCI is an explanatory association index, not a predictor of minimum field size. A SageMath data set contains 52 matroids, including 47 representable cases. Spearman correlation, baseline comparisons, ablation, bootstrap intervals, permutation tests, and leave one family out analyses evaluate the index. Across representable cases, FRCI correlates positively with minimum field size (Spearman $\rho = 0.8527$, $p = 2.81 \times 10^{-14}$). Element count, corank, and cocircuit count show no significant association; rank is negatively associated, and circuit count is only weakly positive. Dependence and coupling contribute most to the observed association. These findings offer a structural account of field size variation relevant to matroidal network coding, while remaining limited to the studied sample.

Keywords: matroids, finite field representability, structural complexity, FRCI

1. Introduction

Matroid theory formalizes the connection between combinatorial independence and linear representation [1]. A matroid is representable over a finite field $GF(q)$ when a matrix over that field has the same independent sets as the matroid. This property also matters for the solvability of linear codes on matroidal networks [2, 3].

Existing work chiefly asks whether a matroid is representable over a specified field or determines the smallest field on which a representation exists. Network coding research has also examined bounds on the required field size [2, 3]. The structural sources of variation in minimum field size across matroids remain less clear. Matroids of comparable size can have different circuits, cocircuits, and local geometry. Conversely, large regular and graphic matroids are representable over $GF(2)$ [1]. Element count alone therefore cannot describe the relevant complexity.

This study investigates how intrinsic structure is associated with minimum field size. It defines the duality invariant V5-MSS signature and an equally weighted FRCI, then evaluates them on 52 representative matroids. The analysis compares simple count based measures, removes individual features, and tests stability

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

with bootstrap, permutation, and leave one family out procedures. The aim is to explain observed variation in field requirements and to clarify where a structure-based index succeeds or fails. FRCI describes association within this data set; it does not decide representability or predict the minimum field size of an arbitrary matroid.

2. Theoretical background

2.1. Matroids and finite field representations

Let $M = (E, I)$ be a finite matroid with ground set E and independent set family I [1]. Write $n = |E|$, $r = r(M)$, and $r = n - r$ for its size, rank, and corank. A circuit is a minimal dependent set. The circuits of the dual matroid M are the cocircuits of M [1]. Finite field representability is preserved under duality, motivating the duality invariant signature used below. The quantity $q_{min}(M)$ is the smallest finite field cardinality for which M has a representation. Field characteristic can impose additional restrictions: the Fano matroid is representable only in characteristic 2, whereas the non Fano matroid requires a characteristic other than 2 [1].

2.2. Matroidal networks and network coding

The existence of a linear solution to a matroidal network is closely tied to finite field representations of its associated matroid [2, 3]. Dougherty, Freiling, and Zeger connected networks, matroids, and non Shannon information inequalities and examined how representability constrains network coding solutions [2, 3]. This connection motivates the analysis of structural factors associated with minimum field size.

2.3. Related work

Three research strands frame this study. Work on representability addresses the fields over which a given matroid can be represented; the resolution of Rota's conjecture belongs to this strand [4]. Excluded minor theory establishes boundaries through Tutte's characterization of binary matroids, Seymour's decomposition of regular matroids, results related to Rota's conjecture, Dowling geometries, and constructions of nonrepresentable matroids [4-8]. Network coding research examines solvability under field size and characteristic constraints [2, 3]. The present analysis instead asks which measured structural properties accompany variation in minimum field size among representative matroids.

3. Methodology

3.1. Field requirements and the structural signature

The field requirement descriptor records minimum field size, the type of field constraint, and representability status:

$$\Psi_q(M) = [q_{min}(M), T_q(M), S_q(M)] \tag{1}$$

Here q_{min} is the minimum finite field cardinality; T_q denotes a constraint class, such as ALL, SIZE ≥ k, CHAR-2, CHAR-NOT-2, CHAR-3, or a nonconsecutive set of admissible fields; and S_q is R or NR. No finite q_{min} is assigned to a nonrepresentable matroid. The duality invariant structural signature is

$$\Phi_{V5}(M) = [\varphi_r(M), \varphi_d(M), \varphi_c(M), \varphi_l(M)] \tag{2}$$

All four components are calculated from the structure of M and take the same values for M and M. They describe related facets rather than statistically independent dimensions, as the feature correlation analysis in Section 5.8 shows.

3.2. Four structural features

The signature combines dual rank imbalance, intrinsic dependence, circuit and cocircuit coupling, and local geometric capacity:

$$\varphi_r(M) = \frac{|n-2r|}{n}, B_C(n,r) = \sum_{k=1}^{min(n,r+1)} \binom{n}{k}, \tag{3}$$

$$\varphi_d(M) = ¼(\eta_C + \eta_{C^*} + s_C + s_{C^*}), \kappa_C(M) = \frac{2}{m(m-1)} \sum_{i<j} J(C_i, C_j), \tag{4}$$

$$\varphi_c(M) = ½(\kappa_C + \kappa_{C^*}), \varphi_l(M) = \frac{q_{LB}(M)-2}{q_{LB}(M)}, q_{LB} = max\{2, \ell - 1, \ell^* - 1\}. \tag{5}$$

In these definitions, η_C is a normalized circuit count, s_C is a normalized mean circuit size, and $J(C_i, C_j)$ is Jaccard overlap between two circuits. The quantity $\ell(M)$ is the maximum number of points in a rank 2 flat. The normalization $B_C(n,r)$ counts subsets of size at most $r+1$ and provides a loose upper benchmark for circuit count; for simple matroids it includes subsets of sizes one and two that cannot be circuits. The analogous quantities for M enter the duality invariant features. The local bound q_{LB} uses both $\ell(M)$ and $\ell(M)$.

Remark 1. Intrinsic dependence φ_d summarizes the abundance and average size of dependent sets, whereas coupling φ_c summarizes overlap among them. They capture different structural aspects but are strongly associated in this data set (Spearman ρ = 0.918; Section 5.8). The ablation analysis examines their separate contributions to FRCI.

3.3. FRCI and equal weighting

FRCI is the arithmetic mean of the four structural features:

$$FRCI(M) = ¼[\varphi_r(M) + \varphi_d(M) + \varphi_c(M) + \varphi_l(M)], FRCI(M) \neq q_{min}(M). \tag{6}$$

Proposition 1 (symmetry of equal weights). Consider a linear aggregate $F = \Sigma_i w_i \varphi_i$ whose weights sum to one. If F is unchanged when any two feature indices are exchanged, then each weight equals one quarter. Because each φ_i is already invariant under matroid duality, duality invariance places no additional restriction on the weights. The uniqueness claim applies to normalized linear aggregates; nonlinear symmetric aggregates, such as geometric means, are outside its scope.

Proof sketch. Permutation symmetry requires all four weights to be equal. Their sum is one, so each equals one quarter. Duality invariance follows from the invariance of each component.

Equal weighting also keeps the four components on a common [0, 1] scale, avoids fitting weights to q_{min} in this sample, and allows additive interpretation. It is a design choice for description rather than an optimization of correlation.

3.4. Local geometric lower bound

Proposition 2 (rank 2 capacity). If a simple matroid M is representable over $GF(q)$, a rank 2 flat contains at most $q+1$ distinct points [1]. Hence $q_{min}(M) \geq \ell(M) - 1$. Applying the same argument to the dual requires its rank 2 flats to be counted as distinct points. With this point based interpretation, $q_{min}(M) \geq q_{LB}(M) = max\{2, \ell(M) - 1, \ell(M) - 1\}$.

Proof sketch. A rank 2 linear representation over $GF(q)$ has $q+1$ one dimensional subspaces, the points of $PG(1,q)$ [1]. Distinct points on a rank 2 flat require distinct subspaces. Dual representability permits the same point count argument for M.

3.5. Computational cost

Circuit and cocircuit enumeration is the main computational bottleneck and has exponential worst case cost [1]. Dependence aggregation is linear in the enumerated circuit count m_C, whereas pairwise coupling is quadratic in m_C. The present matroids have at most 21 elements, and the largest circuit count is approximately 1,172. Larger instances would require sampling, truncation, or incremental enumeration.

4. Experimental design

4.1. Data set

The data set contains 52 representative matroids: 47 representable and five nonrepresentable. The 47 representable cases are grouped by mechanism into Regular/Graphic (12), Uniform (16), Characteristic (9), and Special/SizeBound (10). Relative to an earlier 39 case set, the expanded set adds seven graphic matroids with q_{min} = 2 and six high field uniform matroids U(2, 11) through U(2, 16). SageMath 10.9 generated the matroid structures [9]. The set is purposive rather than a random sample of all matroids.

4.2. Sources of minimum field size and data flow

The q_{min} and T_q labels follow standard representability properties, primarily Oxley [1]. Dowling geometry uses its original construction [7]; the nonrepresentability labels also draw on standard sources [1, 8]. Graphic matroids are regular and have q_{min} = 2 [1]. For $U(2,n)$, q_{min} is the smallest prime power at least $n-1$ [1]. Figure 1 summarizes the generation of matroid structures, their four features, and the statistical analysis. Sample level labels and construction code are retained in the research data files.

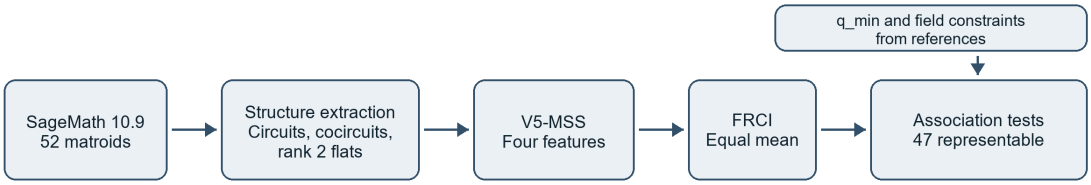


Figure 1. Data generation and structural feature analysis workflow

4.3. Statistical analysis

The evaluation uses Spearman rank correlation [10], simple feature baselines, component ablation, and interfeature correlations. It also computes percentile bootstrap confidence intervals and two sided permutation tests [11], then repeats the main correlation after removing each mechanism family. All reported p values are two sided. The random seed is 20260908. The bootstrap uses 20,000 resamples; the permutation test uses 100,000 permutations. Stratified bootstrap resamples within the four mechanism families.

4.4. Reproducibility

The SageMath construction script, statistical scripts, plotting code, and sample data form the reproducibility materials for this study. They specify the matroid catalog entries, structural calculations, and analysis settings.

5. Results and discussion

5.1. Individual features and FRCI

Table 1 reports the Spearman correlations between each structural feature and q_{min} for the 47 representable matroids.

Table 1. Spearman correlations of V5-MSS features and FRCI with q_{min} ($n = 47$)

Feature	n	Spearman ρ	p value
φ_r Dual rank imbalance	47	0.6163	3.98×10^{-6}
φ_d Intrinsic dependence	47	0.9405	1.02×10^{-22}
φ_c Circuit/cocircuit coupling	47	0.8474	5.88×10^{-14}
φ_l Local geometric capacity	47	0.8153	3.00×10^{-12}
FRCI	47	0.8527	2.81×10^{-14}

All four features have positive associations with q_{min}. Intrinsic dependence φ_d has the strongest single feature correlation ($\rho = 0.941$), followed by circuit and cocircuit coupling φ_c ($\rho = 0.847$). FRCI correlates at $\rho = 0.853$. Figure 2 places these results on the same scale. These observations show that several structural properties accompany field requirements, although the components contribute unequally.

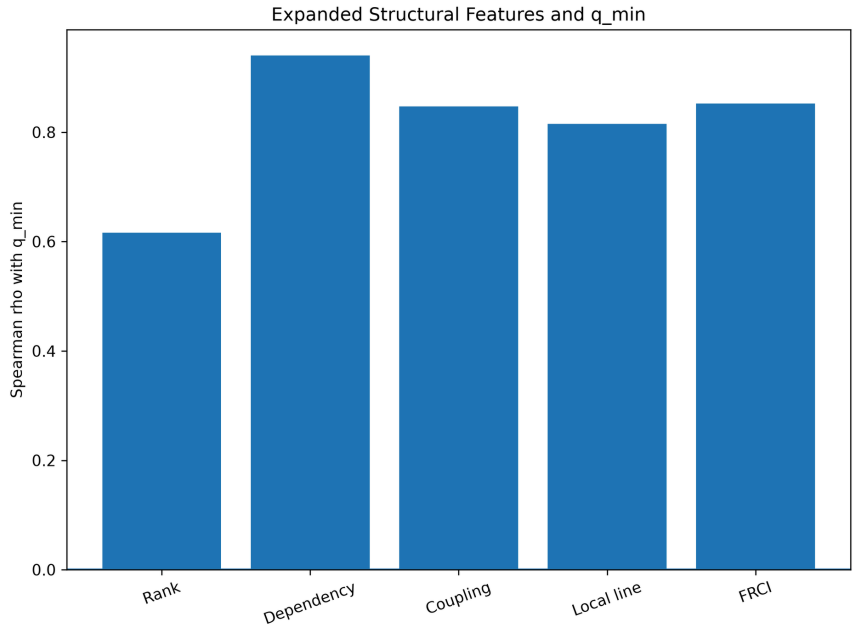


Figure 2. Spearman correlations of individual features and FRCI with q_{min}

5.2. Comparison with simple structural measures

Table 2 compares FRCI with elementary size and count measures.

Table 2. Simple structural baselines for q_{min} ($n = 47$)

Baseline measure	Spearman ρ	p value	Interpretation
Element count n	0.0268	0.858	Not significant
Rank r	−0.6444	1.02×10^{-6}	Negative association
Corank $n - r$	0.2073	0.162	Not significant
Circuit count $\lvert C(M)\rvert$	0.3193	0.0287	Weak positive association
Cocircuit count $\lvert C*(M)\rvert$	−0.2605	0.0769	Not significant
FRCI	0.8527	2.81×10^{-14}	Positive association

Element count and corank have no statistically significant correlation with q_{min} in this data set. Rank has a negative correlation: some large graphic matroids have high rank but $q_{min} = 2$, whereas rank 2 uniform matroids can require larger fields. Circuit count has only a weak positive correlation ($\rho = 0.319$). The multicomponent FRCI therefore captures associations that these simple counts miss, even though the best individual structural component, φ_d, has a higher correlation than FRCI.

5.3. Overall and within family association

Across 47 representable samples, the correlation between FRCI and q_{min} is $\rho = 0.8527$ ($p = 2.81 \times 10^{-14}$). Figure 3 colors the samples by representability mechanism. Table 3 separates the pooled result from within family and sensitivity analyses.

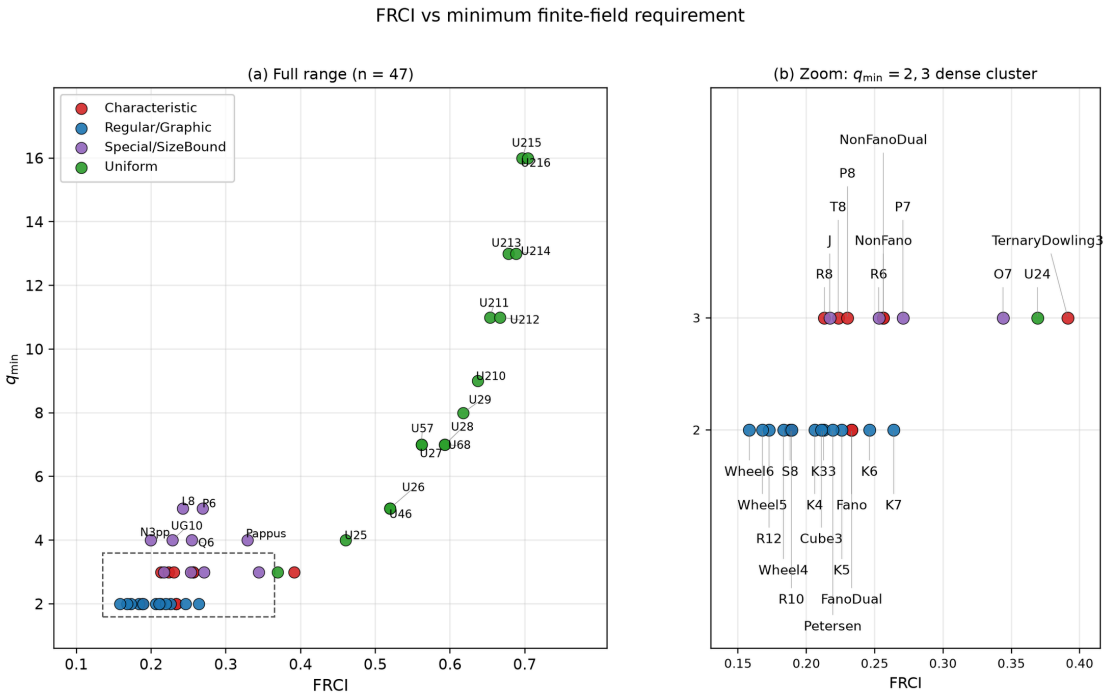


Figure 3. FRCI and minimum finite field size by representability mechanism

Table 3. Correlations by mechanism family and sensitivity subset

Sample group	n	Spearman ρ	p value
All representable	47	0.8527	2.81×10^{-14}
Uniform	16	0.9918	6.31×10^{-14}
Regular/Graphic	12	Undefined	—
Characteristic	9	0.2762	0.472
Special/SizeBound	10	−0.1557	0.667
Non Uniform	31	0.5362	1.87×10^{-3}
Excluding explicit dual controls	42	0.8350	6.26×10^{-12}

The Uniform family is almost monotonic internally (ρ = 0.992). The Regular/Graphic family has constant q_{min} = 2, so its within family correlation is undefined. After all Uniform cases are removed, the remaining 31 cases still have a positive association (ρ = 0.536, p = 0.00187). Removing the explicit duality controls leaves 42 cases and ρ = 0.835. Thus the pooled association is not entirely produced by the Uniform sequence, although cross family differences contribute to its strength. The Characteristic and Special/SizeBound groups do not show a statistically significant positive within family association.

5.4. Component ablation

Table 4 recomputes FRCI after omitting one component at a time and averaging the remaining three.

Table 4. Component ablation of FRCI (n = 47)

Aggregate	Spearman ρ	Change from full
Full FRCI	0.8527	—
Without φ_r	0.9383	+0.086
Without φ_d	0.7678	−0.085
Without φ_c	0.7628	−0.090
Without φ_l	0.8555	+0.003

Removing coupling φ_c reduces the correlation to 0.763, and removing dependence φ_d reduces it to 0.768. Removing rank imbalance φ_r instead raises the correlation to 0.938; removing local capacity φ_l changes it only slightly, to 0.856. Dependence and coupling therefore drive much of the association in this data set. Equal weighting retains all four interpretable structural facets rather than maximizing the sample correlation.

5.5. Bootstrap, permutation, and leave one family out checks

Percentile bootstrap 95% confidence intervals for the main Spearman coefficient are [0.706, 0.939] for ordinary resampling and [0.765, 0.919] for family stratified resampling. The non Uniform subset gives [0.242, 0.746]. All three intervals exclude zero (Figure 4). With 100,000 two sided permutations, the full sample reaches the test's approximate 10^{-5} resolution; the non Uniform sample has $p \approx 0.0019$. These tests describe stability within the selected sample rather than population wide generality.

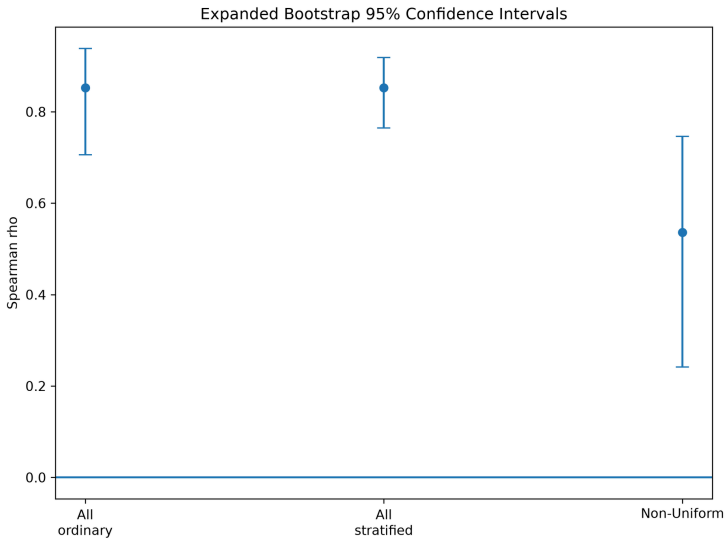


Figure 4. Bootstrap 95% confidence intervals for the FRCI–q_min correlation

Table 5 removes each mechanism family in turn. Every remaining sample has a positive coefficient, ranging from 0.536 to 0.916. Omitting Uniform leaves the weakest association; omitting Special/SizeBound leaves the strongest (Figure 5).

Table 5. Leave one family out sensitivity analysis

Omitted family	Remaining n	Spearman ρ	p value
Characteristic	38	0.8977	2.23×10^{-14}
Regular/Graphic	35	0.8387	3.18×10^{-10}
Special/SizeBound	37	0.9161	1.86×10^{-15}
Uniform	31	0.5362	1.87×10^{-3}

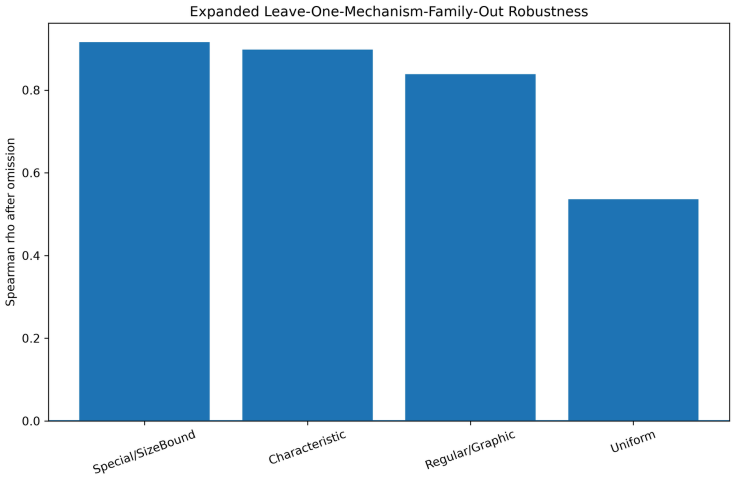


Figure 5. Leave one family out correlations

5.6. Cases with discordant rankings

Figure 6 shows matroids whose q_{min} and FRCI percentile ranks differ most. N3pp (+0.468), L8 (+0.287), and UG10 (+0.277) rank higher by q_{min} than by FRCI. K7 (−0.383) and K6 (−0.277) rank higher by FRCI than by q_{min}. K7 has approximately 1,172 circuits but, as a graphic matroid, remains representable over $GF(2)$. FRCI therefore overstates field demand for some large graphic cases. The local bound q_{LB} equals q_{min} in 27 of the 47 representable samples (57.4%).

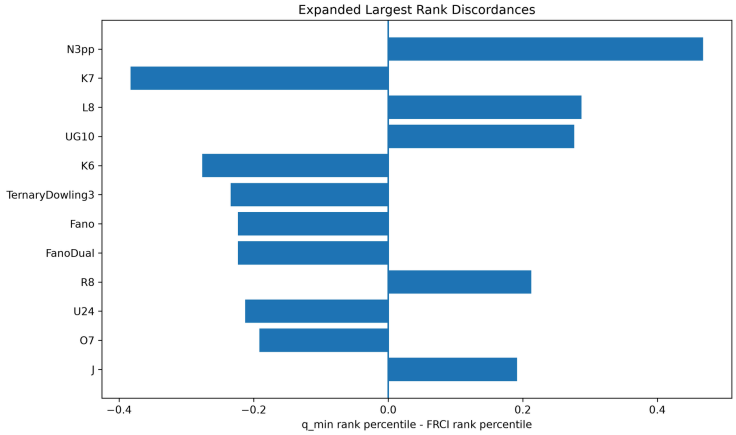


Figure 6. Largest discrepancies between q_{min} and FRCI percentile ranks

5.7. Duality and non Uniform cases

The five explicit dual pairs have zero absolute difference in every signature component and in FRCI. This sample level check confirms the intended duality invariance of the implementation. Figure 7 plots the non Uniform representable cases using intrinsic normalization and distinguishes the original and expanded sets. Across the full data set, FRCI ranges overlap between nonrepresentable cases (0.203–0.358) and representable cases (0.158–0.704). FRCI cannot classify representability.

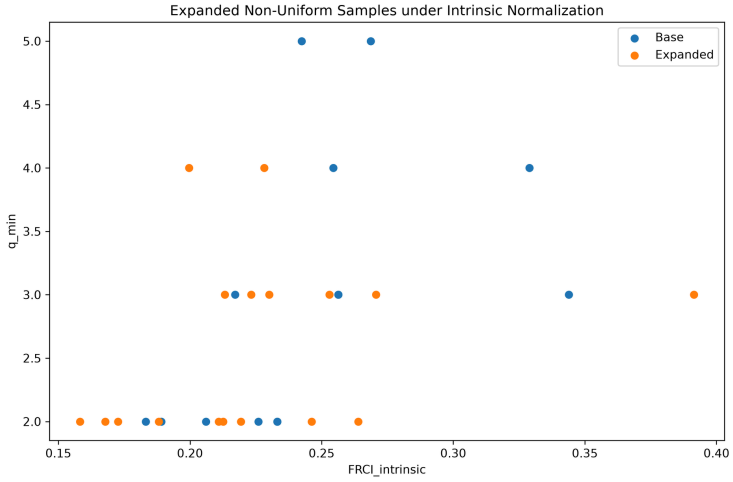


Figure 7. Non Uniform cases under intrinsic normalization

5.8. Correlations among the features

Table 6 and Figure 8 show the Spearman correlation matrix for the four features, FRCI, and q_{min}.

Table 6. Spearman correlation matrix for structural features, FRCI, and q_{min} ($n = 47$)

Feature	φ_r	φ_d	φ_c	φ_l	FRCI	q_{min}
φ_r	1.000	0.605	0.447	0.810	0.837	0.616
φ_d	0.605	1.000	0.918	0.841	0.893	0.941
φ_c	0.447	0.918	1.000	0.746	0.793	0.847
φ_l	0.810	0.841	0.746	1.000	0.875	0.815
FRCI	0.837	0.893	0.793	0.875	1.000	0.853
q_{min}	0.616	0.941	0.847	0.815	0.853	1.000

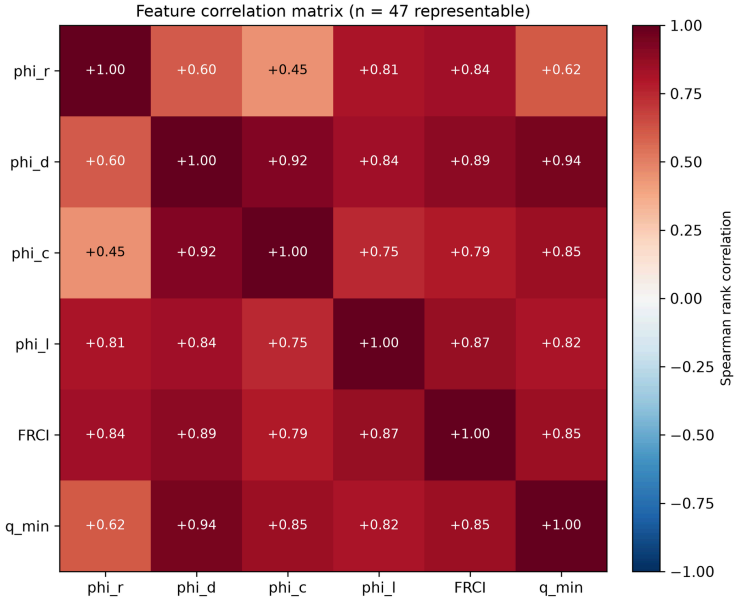


Figure 8. Spearman correlation matrix of features, FRCI, and q_{min}

Dependence φ_d and coupling φ_c have the strongest interfeature Spearman correlation (0.918; Pearson correlation 0.986). They are still distinguishable: their correlations with q_{min} are 0.941 and 0.847, respectively, and removing either weakens the three component aggregate. Rank imbalance φ_r has lower correlations with some other features (as low as 0.447); local capacity φ_l contributes a separate geometric interpretation. The four features should therefore be read as associated facets rather than orthogonal dimensions. Their correlation does not invalidate FRCI as a descriptive aggregate, but it limits any interpretation of each component as an independent causal effect.

5.9. Interpretation and limits

The measured dependence and coupling features summarize constraints among circuits and cocircuits. Local rank 2 capacity supplies a combinatorial field size bound under the point based interpretation of Proposition 2. Rank imbalance captures a different global aspect of structure. Together, these mechanisms help explain why some matroids with similar element counts have different q_{min} values.

The index works best when local geometric capacity or dependence structure tracks the field requirement, as in parts of the Uniform family. Field characteristic restrictions, nonconsecutive admissible fields such as those of the Pappus matroid, and global representation equations can instead dominate q_{min}. Graphic matroids K6 and K7 illustrate another limit: many circuits can raise FRCI even though regularity guarantees a $GF(2)$ representation.

Several findings constrain the interpretation. The non Uniform correlation is lower than the pooled value (0.536 versus 0.853), so much of the pooled pattern reflects the Uniform series and differences among mechanism families. FRCI does not outperform φ_d as a single feature in rank correlation (0.853 versus 0.941). Its value lies in duality invariance and additive description. Representable and nonrepresentable samples overlap in FRCI, and full circuit enumeration is costly on large matroids. Finally, the catalog and systematically constructed examples were selected deliberately. Their p values quantify association and stability for this set; they do not establish a universal law for all matroids.

6. Conclusion

This study examined the association between intrinsic matroid structure and minimum finite field size. It defined a duality invariant four component signature, V5-MSS, and its equally weighted descriptive index, FRCI. The analysis also established the symmetry rationale for equal weights and a local geometric field size bound.

Among 47 representable cases within a 52 matroid data set, FRCI correlated positively with minimum field size (Spearman ρ = 0.8527). Element count, corank, and cocircuit count showed no significant association, while rank correlated negatively and circuit count only weakly positively. Ablation and feature correlation analyses identified dependence and circuit and cocircuit coupling as the main contributors. Bootstrap, permutation, and leave one family out checks supported a positive association in the selected data set.

The evidence also sets clear limits. The pooled correlation weakens after Uniform cases are removed, and one dependence feature correlates more strongly with minimum field size than the aggregate. Characteristic constraints and global representability equations remain difficult to describe with the four features. Larger and more varied matroid samples, efficient approximations to circuit enumeration, and tests on concrete matroidal network coding problems are natural next steps. FRCI should be used to interpret structural associations, not to predict q_{min} or determine representability.

References

[1] Oxley, J. G. (2011). *Matroid theory* (2nd ed.). Oxford University Press.

[2] Dougherty, R., Freiling, C., & Zeger, K. (2007). Networks, matroids, and non-Shannon information inequalities. *IEEE Transactions on Information Theory, 53*(6), 1949–1969.

[3] Dougherty, R., Freiling, C., & Zeger, K. (2011). Network coding and matroid theory. *Proceedings of the IEEE, 99*(3), 388–405.

[4] Geelen, J., Gerards, B., & Whittle, G. (2014). Solving Rota's conjecture. *Notices of the American Mathematical Society, 61*(7), 736–743.

[5] Tutte, W. T. (1958). A homotopy theorem for matroids, I. *Transactions of the American Mathematical Society, 88*(1), 144–160.

[6] Seymour, P. D. (1980). Decomposition of regular matroids. *Journal of Combinatorial Theory, Series B, 28*(3), 305–359.

[7] Dowling, T. A. (1973). A class of geometric lattices based on finite groups. *Journal of Combinatorial Theory, Series B, 14*(1), 61–86.

[8] Vámos, P. (1978). The missing axiom of matroid theory is lost forever. *Journal of the London Mathematical Society, s2-18*(3), 403–408.

[9] The Sage Developers. (2026). *SageMath, the Sage Mathematics Software System, Version 10.9*. https://www.sagemath.org

[10] Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology, 15*(1), 72–101.

[11] Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman & Hall.